

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:12:45 ; Search time 36 Seconds

(without alignments)

636.642 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 172

Sequence: 1 MMLHIQIAAAALSVLTFM.....IAFGTNEAMNSQNRRAELSY 172

Scoring table:

OLIGO Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	172	100.0	172 21	AAV55089
2	107	62.2	172 21	AAV55080
3	107	62.2	172 21	AAV55091
4	68	39.5	172 21	AAV55092
5	15	8.7	16 21	AAV55093
6	15	8.7	16 21	AAV55094
7	14	8.1	14 21	AAV55094
8	14	8.1	14 21	AAV55094
9	9	5.2	2168 22	ABR64563
10	8	4.7	139 22	ABG22196

11	8	4.7	158	21	AAV55097	Zea mays protein f
12	8	4.7	306	22	ABR62221	Drosophila melanog
13	8	4.7	436	22	ABG11084	Novel human diagno
14	8	4.7	452	21	AAV74388	Neisseria meningit
15	8	4.7	545	22	ABR58578	Drosophila melanog
16	8	4.7	545	22	ABR67264	Drosophila melanog
17	8	4.7	606	23	AAV75084	Ryegrass 4-coumar
18	8	4.7	637	22	ABR70398	Drosophila melanog
19	8	4.7	637	22	ABR66879	Drosophila melanog
20	8	4.7	644	23	AAV10024	Rice acetolactica
21	8	4.7	644	23	AAV10024	Rice herbicide res
22	8	4.7	662	22	ABR59430	Drosophila melanog
23	8	4.7	1164	22	ABR57802	Drosophila melanog
24	8	4.7	1412	22	ABR60943	Drosophila melanog
25	8	4.7	2090	22	ABR64682	Drosophila melanog
26	8	4.7	2703	22	ABR63299	Drosophila melanog
27	8	4.7	9	18	AAV64637	Null peptide that
28	8	4.1	10	19	AAV61561	Biotinylated pepti
29	8	4.1	13	16	AAV84046	Murine MHC class I
30	8	4.1	14	17	AAV02569	AKAP79 A37-50 muta
31	8	4.1	14	21	AAV14909	Mutant peptide AKA
32	8	4.1	15	16	AAV84047	Murine MHC class I
33	8	4.1	21	22	ABR37792	Human peptide #593
34	8	4.1	21	22	AAV53912	Peptide #5276 enco
35	8	4.1	21	22	AAV53912	Human brain expres
36	8	4.1	21	22	AAV18708	Human bone marrow
37	8	4.1	21	22	AAV18708	Peptide #5142 enco
38	8	4.1	21	22	AAV18708	Peptide #614 enco
39	8	4.1	21	22	AAV18708	Human peptide enco
40	8	4.1	21	22	AAV18708	Human RIZ alternat
41	8	4.1	21	22	AAV18708	5' terminus of hum
42	8	4.1	21	22	AAV18708	Human Rb-interacti
43	8	4.1	21	22	AAV18708	HIV tat related am
44	8	4.1	21	22	AAV18708	Arctic fish antifr
45	8	4.1	21	22	AAV18708	Antifreeze protein
46	8	4.1	21	22	AAV18708	Protein derived fr
47	8	4.1	21	22	AAV18708	P. americanus anti
48	8	4.1	21	22	AAV18708	P. americanus anti
49	8	4.1	21	22	AAV18708	Winter flounder il
50	8	4.1	21	22	AAV18708	Melting pt. depres
51	8	4.1	21	22	AAV18708	P. americanus anti
52	8	4.1	21	22	AAV18708	Synthetic antifr
53	8	4.1	21	22	AAV18708	Synthetic antifr
54	8	4.1	21	22	AAV18708	Synthetic antifr
55	8	4.1	21	22	AAV18708	Synthetic antifr
56	8	4.1	21	22	AAV18708	Human EST encoded
57	8	4.1	21	22	AAV18708	Synthetic antifr
58	8	4.1	21	22	AAV18708	Synthetic antifr
59	8	4.1	21	22	AAV18708	Synthetic antifr
60	8	4.1	21	22	AAV18708	Synthetic antifr
61	8	4.1	21	22	AAV18708	Synthetic antifr
62	8	4.1	21	22	AAV18708	Synthetic antifr
63	8	4.1	21	22	AAV18708	Synthetic antifr
64	8	4.1	21	22	AAV18708	Synthetic antifr
65	8	4.1	21	22	AAV18708	Synthetic antifr
66	8	4.1	21	22	AAV18708	Synthetic antifr
67	8	4.1	21	22	AAV18708	Synthetic antifr
68	8	4.1	21	22	AAV18708	Synthetic antifr
69	8	4.1	21	22	AAV18708	Synthetic antifr
70	8	4.1	21	22	AAV18708	Synthetic antifr
71	8	4.1	21	22	AAV18708	Synthetic antifr
72	8	4.1	21	22	AAV18708	Synthetic antifr
73	8	4.1	21	22	AAV18708	Synthetic antifr
74	8	4.1	21	22	AAV18708	Synthetic antifr
75	8	4.1	21	22	AAV18708	Synthetic antifr
76	8	4.1	21	22	AAV18708	Synthetic antifr
77	8	4.1	21	22	AAV18708	Synthetic antifr
78	8	4.1	21	22	AAV18708	Synthetic antifr
79	8	4.1	21	22	AAV18708	Synthetic antifr
80	8	4.1	21	22	AAV18708	Synthetic antifr
81	8	4.1	21	22	AAV18708	Synthetic antifr
82	8	4.1	21	22	AAV18708	Synthetic antifr
83	8	4.1	21	22	AAV18708	Synthetic antifr

84	7	4.1	75	22	AAG62224	Human gene 24-enco
85	7	4.1	75	23	ABG63540	Human albumin fusi
86	7	4.1	75	23	ABG63541	Human albumin fusi
87	7	4.1	78	22	AAE09778	Adenovirus hexon
88	7	4.1	79	22	AAU86705	Novel human connec
89	7	4.1	82	20	AAZ23877	Winter flounder an
90	7	4.1	88	21	AA887224	Human signal pepti
91	7	4.1	97	20	AA886160	P. americanus anti
92	7	4.1	102	21	AA632715	Zea mays protein f
93	7	4.1	106	22	ABG03609	Novel human diagno
94	7	4.1	109	22	AA865761	Cysteine protease
95	7	4.1	119	22	AAU94079	Human reproductive
96	7	4.1	121	23	AAU98749	Chicken anemia vir
97	7	4.1	125	21	AA657848	Zea mays protein f
98	7	4.1	131	20	AAW23876	A hypothetical mat
99	7	4.1	131	20	AAW95195	Hypothetical hybri
100	7	4.1	131	21	AAV92493	TILCV C3 mutant, m

ALIGNMENTS

RESULT 1

AAV55089

ID AAV55089 standard; Protein: 172 AA.

AAV55089;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #1.

BASB019 protein; diagnosis; infectious organism; auditive nerve damage; genetic mutation screening; antibody production; vaccine; otitis media; bacterial infection; pneumonia; sinusitis; nosocomial infection; invasive disease; delayed speech learning; bacteria adhesion prevention; upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

WO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMIRK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AAZ40351.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 10pp; English.

This sequence is a Moraxella catarrhalis BASB019 protein of the invention. The sequences can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification. Identification of mutation in BASB019 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies, and as a target for the screening of antimicrobial drugs. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases including bacterial infection, otitis media in infants and

CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

Sequence 172 AA;

Query Match 100.0%; Score 172; DB 21; Length 172;
Best Local Similarity 100.0%; Pred. No. 1, 6e-154; Indels 0; Gaps 0;
Matches 172; Conservative 0; Mismatches 0;

QY	1	MMHIIQIAAAALSVLTFTMGCAKSTSOVMAAPNAPGYGVYTGVAPLVNDERYK	60
DB	1	MMHIIQIAAAALSVLTFTMGCAKSTSOVMAAPNAPGYGVYTGVAPLVNDERYK	60
QY	61	ALASKLPFLVYFDFPSDEIKPQAAALDEQAQFLTTNOTARVAVAGHTDERGSRREYNMSL	120
DB	61	ALASKLPFLVYFDFPSDEIKPQAAALDEQAQFLTTNOTARVAVAGHTDERGSRREYNMSL	120
QY	121	GERRAVAVRNITLKGINQASVEIISGEEPRIFGTNEAMSQNRRAELSY	172
DB	121	GERRAVAVRNITLKGINQASVEIISGEEPRIFGTNEAMSQNRRAELSY	172

RESULT 2

AAV55090

ID AAV55090 standard; Protein: 172 AA.

AAV55090;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #2.

BASB019 protein; diagnosis; infectious organism; auditive nerve damage; genetic mutation screening; antibody production; vaccine; otitis media; bacterial infection; pneumonia; sinusitis; nosocomial infection; invasive disease; delayed speech learning; bacteria adhesion prevention; upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

WO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMIRK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AAZ40352.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 10pp; English.

This sequence is a Moraxella catarrhalis BASB019 protein of the

XX 06-MAY-1998; 98GB-0009683.
 PR (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PA Ruelle J;
 XX WPI: 2000-062148/05.
 DR N-PSDB; AA440354.
 XX
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarhalis used to prepare vaccines against bacterial infections
 PS Claim 3; Fig 3; 101pp; English.

XX This sequence is a Moraxella catarhalis BASB019 protein of the
 CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a
 CC target for the screening of antimicrobial drugs. The polypeptides can
 CC also be used in vaccine formulations, and to identify agonists and
 CC antagonists. The polypeptides, antibodies, agonists and antagonists
 CC (which are bacteriostatic) are used for the treatment and prevention of
 CC diseases including bacterial infection, otitis media in infants and
 CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed speech
 CC learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.

XX Sequence 172 AA;
 SQ
 Query Match 39.5%; Score 68; DB 21; Length 172;
 Best Local Similarity 100.0%; Pred. No. 3.2e-56;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 LPSLVYFDPDSDEIRKQAAATIDEOAFITNOTARVAVAGHTDEKSGREYNMSLGRRA 125
 |||||||
 DB 66 LPSLVYFDPDSDEIRKQAAATIDEOAFITNOTARVAVAGHTDEKSGREYNMSLGRRA 125
 |||||||

QY 126 VAVRNYTL 133
 |||||||
 DB 126 VAVRNYTL 133
 |||||||

RESULT 5
 AAY55093
 ID AAY55093 standard; Protein: 16 AA.
 XX
 AC AAY55093;
 XX
 DT 01-MAR-2000 (first entry)
 XX
 DE M. catarhalis BASB019 protein sequence fragment.
 XX
 KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KW genetic mutation screening; antibody production; vaccine; otitis media;
 KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; delayed speech learning; bacteria adhesion prevention;
 KW upper respiratory tract infection; middle ear infection; therapy.

XX Moraxella catarhalis.
 OS
 XX
 PN W09957277-A2.
 PD 11-NOV-1999.
 XX
 XX 03-MAY-1999; 99MO-EP03038.
 PF
 PR 06-MAY-1998; 98GB-0009683.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PL Ruelle J;
 XX
 DR WPI: 2000-062148/05.
 XX
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarhalis used to prepare vaccines against bacterial infections
 PS Example 8; Page 70; 101pp; English.

XX This sequence is a fragment of a Moraxella catarhalis BASB019 protein of
 CC the invention. The sequences can be used for diagnosis of disease,
 CC staging of disease, or determining response of an infectious organism to
 CC drugs. The polynucleotides may be used as a source for hybridisation
 CC probes, and for screening of genetic mutations, serotype, organism or
 CC strain identification, identification of mutation in BASB013 sequences,
 CC and as components of arrays which are useful for diagnostic and
 CC prognostic purposes. The polypeptides can be used to produce antibodies,
 CC and as a target for the screening of antimicrobial drugs. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC agonists and antagonists. The polypeptides, antibodies, agonists and
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases including bacterial infection, otitis media in
 CC infants and children, pneumonia in the elderly, sinusitis, nosocomial
 CC infections and invasive diseases, chronic otitis media with hearing loss,
 CC fluid accumulation in the middle ear, auditive nerve damage, delayed
 CC speech learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.

XX Sequence 16 AA;
 SQ
 Query Match 8.7%; Score 15; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 4.8e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 NEAWSONRRAELSY 172
 |||||||
 DB 2 NEAWSONRRAELSY 16
 |||||||

RESULT 6
 AAY57577
 ID AAY57577 standard; peptide: 16 AA.
 XX
 AC AAY57577;
 XX
 DT 02-MAR-2000 (first entry)
 XX
 DE Moraxella catarhalis BASB020 specific peptide SEQ ID NO:13.
 XX
 KW Moraxella catarhalis; diagnosis; vaccine; infection;
 KW otitis media; pneumonia; sinusitis; anti-inflammatory; auditory;

KW antibacterial; immune response; immunisation; vaccine.
 XX
 OS Synthetic.
 OS Moraxella catarrhalis.
 PN WO958684-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 07-MAY-1999; 99WO-EP03257.
 XX
 PR 13-MAY-1998; 98GB-0010285.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Rhonard J;
 DR WPI; 2000-062301/05.
 XX
 PT Novel peptides useful as vaccines for Moraxella infections such as
 PT otitis media, pneumonia, sinusitis etc.,
 XX
 PS Example 9; Page 69; 113pp; English.
 XX
 CC The present sequence represents a peptide specific for a BASB020
 CC protein, isolated from Moraxella catarrhalis, used in an example from
 CC the present invention. BASB020 polynucleotide fragments may be used as
 CC hybridisation probes for RNA, cDNA and genomic DNA to isolate full-length
 CC cDNAs and genomic clones encoding BASB020 and to isolate cDNA and genomic
 CC clones of other genes that have high sequence identity to BASB020 gene.
 CC The BASB020 polynucleotides and polypeptides are used as research
 CC reagents and materials for discovery of treatments of and diagnostics for
 CC human diseases. Probes comprising BASB020 nucleotide sequences can be
 CC constructed to conduct efficient screening of genetic mutations,
 CC serotype, taxonomic classification or identification. The polynucleotide
 CC sequences can be used in the discovery and development of antibacterial
 CC compounds. The polypeptides and polynucleotides are used to block the
 CC initial physical interaction between a gram negative and/or gram positive
 CC bacteria to the mammalian host. The polynucleotides encoding certain
 CC non-variable regions of bacterial cell surface protein are used in
 CC polynucleotide constructs which are useful for genetic immunisation
 CC experiments in animal models of infection with M.catarrhalis to
 CC identify protein epitopes able to provoke a prophylactic or therapeutic
 CC immune response. A therapeutic composition comprising an antibody
 CC directed against BASB020 can be used for treating humans with
 CC M. catarrhalis diseases such as sinusitis, otitis media and nosocomial
 CC infections.
 CC
 XX Sequence 16 AA:
 SQ
 Query Match 8.7%; Score 15; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 4.8e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 158 NEEAWSQNRRAELSY 172
 DB 2 NEEAWSQNRRAELSY 16
 RESULT 7
 AAY55094
 ID AAY55094 standard; Protein; 14 AA.
 XX
 AC AAY55094;
 XX
 DT 01-MAR-2000 (first entry)
 XX
 DE M. catarrhalis BASB019 protein sequence fragment.
 XX
 KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KW genetic mutation screening; antibody production; vaccine; otitis media;
 KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; delayed speech learning; bacteria adhesion prevention;

KW upper respiratory tract infection; middle ear infection; therapy.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO957277-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 03-MAY-1999; 99WO-EP03038.
 XX
 PR 06-MAY-1998; 98GB-0009683.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Ruelle J;
 DR WPI; 2000-062148/05.
 XX
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 PT catarrhalis used to prepare vaccines against bacterial infections
 XX
 PS Example 8; Page 70; 101pp; English.
 XX
 CC This sequence is a fragment of a Moraxella catarrhalis BASB019 protein of
 CC the invention. The sequences can be used for diagnosis of disease,
 CC staging of disease, or determining response of an infectious organism to
 CC drugs. The polynucleotides may be used as a source for hybridisation
 CC probes, and for screening of genetic mutations, serotype, organism or
 CC strain identification, identification of mutation in BASB013 sequences,
 CC and as components of arrays which are useful for diagnostic and
 CC prognostic purposes. The polypeptides can be used to produce antibodies,
 CC and as a target for the screening of antimicrobial drugs. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC agonists and antagonists. The polypeptides, antibodies, agonists and
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases including bacterial infection, otitis media in
 CC infants and children, pneumonia in the elderly, sinusitis, nosocomial
 CC infections and invasive diseases, chronic otitis media with hearing loss,
 CC fluid accumulation in the middle ear, auditive nerve damage, delayed
 CC speech learning, infection of the upper respiratory tract and bacteria
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used screen for new antibacterial compounds that may target these
 CC resistant bacteria.
 CC
 XX Sequence 14 AA:
 SQ
 Query Match 8.1%; Score 14; DB 21; Length 14;
 Best Local Similarity 100.0%; Pred. No. 3.7e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 46 YTGVAPLVDNDDEV 59
 DB 1 YTGVAPLVDNDDEV 14
 RESULT 8
 AAY57578
 ID AAY57578 standard; peptide; 14 AA.
 XX
 AC AAY57578;
 XX
 DT 02-MAR-2000 (first entry)
 XX
 DE Moraxella catarrhalis BASB020 specific peptide SEQ ID NO:14.
 XX
 KW Moraxella catarrhalis; BASB020; diagnosis; vaccine; infection;

[illegible]

OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
XX	
PA	11-JUL-2000; 2000US-0614150.
XX	
PI	(PEKE) PE CORP NY.
XX	
PB	Venter JC, Adams M, Li PWD, Myers EW;
XX	
PI	WPI: 2001-656860/75.
DR	N-PSTDB; ABL08666.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX	
PS	Disclosure: SEQ ID NO 20481; 21bp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA sequences (AB01840-ABL16175) and the encoded proteins (AAB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
CC	
CC	
XX	
SQ	Sequence 2168 AA;
OY	8 AAAAATLSV 16
Db	979 AAAAATLSV 987
RESULT 10	
ABG22196	
ID	ABG22196 standard; Protein; 139 AA.
XX	
AC	ABG22196;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #22187.
XX	
KW	Human; chromosome mapping; gene mapping; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
XX	
PA	23-AUG-2000; 2000US-0649167.
XX	
PI	(HYSE-) HYSED INC.
XX	

PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR N-PSDB; AAS86383.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX
XX Claim 20; SEQ ID No 52555; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 139 AA:
SO
Query Match 4.7%; Score 8; DB 22; Length 139;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 AAAAAAALS 15
Db 14 AAAAAAALS 21
RESULT 11
AAG18977
ID AAG18977 standard; Protein; 158 AA.
XX
AC AAG18977;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 20601.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
XX Zea mays subsp. mays.
OS
XX
XX EPI033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135533.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 11-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141267.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.

PS Disclosure: SEQ ID NO 13455; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AAB57737-AB12072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 306 AA:

Query Match 4.7%; Score 8; DB 22; Length 306;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAAAAAL 14
|||||||
Db 195 TAAAAAAL 202

RESULT 13
ABG11084
ID ABG11084 standard; Protein: 436 AA.

AC ABG11084;
XX
XX 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #11075.

XX
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.

OS
XX
XX WO200175067-A2.

PN
XX
XX 11-OCT-2001.

PD
XX
XX 30-MAR-2001; 2001WO-US08631.

PE
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

XX
XX (HSE-) HXSEQ INC.

PA
XX
XX Drmanac RT, Liu C, Tang YF;
PI
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS75271.

XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX
XX
XX Claim 20; SEQ ID NO 41443; 103pp; English.

XX
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 436 AA:

Query Match 4.7%; Score 8; DB 22; Length 436;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAAL 15
|||||||
Db 13 AAAAAAAL 20

RESULT 14
AAY74388
ID AAY74388 standard; Protein: 452 AA.

AC AAY74388;
XX
XX 21-MAR-2000 (first entry)

DE *Neisseria meningitidis* ORF 081 protein sequence SEQ ID NO:252.

XX
XX
XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
KW antigenic; diagnostic; immunogenic; infection; meningitis; septicaemia;
XX antibacterial; gene therapy.

XX
XX
XX *Neisseria meningitidis*.

OS
XX
XX WO957280-A2.

PN
XX
XX 11-NOV-1999.

PD
XX
XX 30-APR-1999; 99WO-US09346.

PE
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.

XX
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

XX
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,
PI Petersen J, Pizzo M, Rappoli R, Ratti G, Scalato E, Scarselli M,
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR N-PSDB; AAZ53150.

XX
XX
XX Novel *Neisseria* polypeptides predicted to be useful antigens for
PT vaccines and diagnostics.

XX
XX
XX Claim 2; Page 267; 1453pp; English.

XX
XX
XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 452 AA;

Query Match 4.7%; Score 8; DB 21; Length 452;
Best Local Similarity 100.0%; Pred. No. 39;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
| | | | | | | |

Db 289 AAAAAAALS 296

RESULT 15

ABBS8578
ID ABB58578 standard; Protein: 545 AA.

XX ABB58578;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 2526.

KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL02681.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 2526; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 545 AA;

Query Match 4.7%; Score 8; DB 22; Length 545;

Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
| | | | | | | |

Db 369 AAAAAAALS 376

RESULT 16

ABB67264
ID ABB67264 standard; Protein: 545 AA.

XX ABB67264;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 28584.

KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL11367.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 28584; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 545 AA;

Query Match 4.7%; Score 8; DB 22; Length 545;
Best Local Similarity 100.0%; Pred. No. 46;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
| | | | | | | |

Db 369 AAAAAAALS 376

RESULT 17

AAU75084
ID AAU75084 standard; Protein: 606 AA.

XX AAU75084;

XX 23-APR-2002 (first entry)
DT Ryegrass 4-coumarate Co-A-lyase 1 (LP4CL1) protein.
DE
XX
KM Perennial ryegrass; lignin; 4 coumarate CoA-lyase; QTL;
XX lignin biosynthesis; enzyme; cinnamoyl-CoA reductase; CCR;
KM cinnamyl alcohol dehydrogenase; CAD; molecular genetic marker;
KM qualitative trait loci; tagging; QTL mapping; DNA fingerprinting;
KM marker assisted selection; forage improvement; turf grass improvement;
KM dry matter digestibility; herbage quality; palatability; regrowth;
XX cold tolerance; drought tolerance; tiller survival; plant persistence.
OS
XX Lolium perenne.
FH Location/Qualifiers
FT MISC-difference 570..571
FT /note="Encoded by GCCTGAAGA"
XX
XX MO200195702-A1.
XX
XX 20-DEC-2001.
XX
XX 14-JUN-2001; 2001WO-AU00699.
XX
XX 14-JUN-2000; 2000AU-0008154.
XX
XX (VICT-) STATE VICTORIA DEPT NATURAL RES & ENVIRO.
XX (UYAD-) UNIV ADELAIDE.
XX (ITMA-) INT MAIZE & WHEAT IMPROVEMENT CENT.
XX (SACS-) STATE SOUTH AUSTRALIA SOUTH AUSTRALIAN R.
XX (UYSC-) UNIV SOUTHERN CROSS.
XX (DAIR-) DAIRY RES & DEV CORP.
XX
XX Spangenberg GC, Lidgett AJ, Heath RL, McInnes RL, Lynch DP;
XX WPI; 2002-097993/13.
XX DR N-PSDB; ABR13751.
XX
XX Novel nucleic acid encoding enzymes involved in lignin biosynthetic
XX pathway from ryegrass or fescue species useful for modifying lignin
XX biosynthesis in plants and as a molecular genetic marker -
XX
XX Claim 13; Fig 2; 148pp; English.
XX
XX This invention represents purified or isolated nucleic acid and protein
XX sequences of enzymes involved in lignin biosynthesis. The enzymes
XX of the invention are 4 coumarate CoA-lyase (4CL), cinnamoyl-CoA
XX reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) from a ryegrass
XX (Lolium sp.) or fescue (Festuca sp.). The invention also comprises an
XX isolated regulatory element from the nucleic acid sequences and a plant
XX cell or seed transformed with the nucleic acid. An isolated regulatory
XX element from these nucleotide molecules is useful for expressing an
XX exogenous gene in plant cells. The nucleotide sequences of the invention
XX and vectors containing these sequences are useful for modifying lignin
XX biosynthesis in a plant and are useful as a molecular genetic marker for
XX qualitative trait loci (QTL) tagging, QTL mapping, DNA fingerprinting
XX and in marker assisted selection, in forage and turf grass improvement,
XX e.g. tagging QTLs for dry matter digestibility, herbage quality,
XX palatability, regrowth after cutting and grazing, cold tolerance,
XX drought tolerance, tiller survival and plant persistence. The present
XX sequence represents the perennial ryegrass 4-coumarate Co-A-lyase 1
XX (LP4CL1) protein of the invention.
XX
XX Sequence .606 AA;
SQ

Query Match 4.7%; Score 8; DB 23; Length 606;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 QIAAAAAA 13
Db 13 QIAAAAAA 20

RESULT 18
ID ABB70398 standard; Protein; 624 AA.
XX ABB70398;
XX AC ABB70398;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 37986.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL14501.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 37986; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 624 AA;
SQ

Query Match 4.7%; Score 8; DB 22; Length 624;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 AAAAAAALS 15
Db 158 AAAAAAALS 165

RESULT 19
ID ABB66879
XX ABB66879 standard; Protein; 637 AA.
XX
XX ABB66879;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 27429.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI: 2001-656860/75.
 N-PSDB; ABL10982.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 XX Disclosure; SEQ ID NO 27429; 21bp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABU10840-ABU16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX
 XX Sequence 637 AA;
 SQ
 Query Match 4.7%; Score 8; DB 22; Length 637;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 AAAAALS 15
 |||||
 DB 461 AAAAALS 468
 RESULT 20
 ABB81482
 ID ABB81482 standard; Protein; 644 AA.
 XX
 XX ABB81482;
 AC
 XX 30-AUG-2002 (first entry)
 DT
 XX Rice acetolactate acid synthase protein SEQ ID NO:1.
 DE
 XX Rice; Oryza sativa var. Kinmaze; acetolactate acid synthase; enzyme;
 KM herbicide resistance; pyrimidinylcarboxy-based herbicide; plant.
 XX
 XX Oryza sativa.
 OS
 XX WO200244385-A1.
 PN
 XX 06-JUN-2002.
 PD
 XX 16-NOV-2001; 2001WO-JP10014.
 PF
 XX 29-NOV-2000; 2000JP-0362630.
 PR
 XX (TSUB) KUMIAT CHEM IND CO LTD.

PA (NAG-) NAT INST AGROBIOLOGICAL SCI.
 PI Shimizu T, Nakayama I, Nagayama K, Fukuda A, Tanaka Y, Raku K;
 XX WPI: 2002-490301/52.
 DR N-PSDB; ABN89399.
 DR
 XX Gene encoding acetolactate acid synthase, useful in providing new breeds
 PT of plants with high resistance against pyrimidinylcarboxy-based
 PT herbicides -
 PS
 XX Claim 1; Page 82-86; 96pp; Japanese.
 XX
 XX The present sequence represents acetolactate acid synthase (I) isolated
 CC from Oryza sativa var. Kinmaze (rice). (I) has resistance against
 CC pyrimidinylcarboxy (PC)-based herbicides as well as acetolactate acid
 CC synthase activity (I) can be used for providing plants with high
 CC resistance against PC-based herbicides.
 CC
 XX Sequence 644 AA;
 SQ
 Query Match 4.7%; Score 8; DB 23; Length 644;
 Best Local Similarity 100.0%; Pred. No. 53;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 AAAAALS 15
 |||||
 DB 6 AAAAALS 13
 RESULT 21
 AAU10024
 ID AAU10024 standard; Protein; 644 AA.
 XX
 XX AAU10024;
 AC
 XX 08-MAY-2002 (first entry)
 DT
 XX Rice acetohydroxyacid synthase (AHS) wild type protein sequence.
 DE
 XX Rice; AHS; acetohydroxyacid synthase; sulphonylurea herbicide;
 KM herbicide resistance; weed control; imidazolinone; EC.4.1.3.18.
 XX
 XX Oryza sativa.
 OS
 XX WO200185970-A2.
 PN
 XX 15-NOV-2001.
 PD
 XX 09-MAY-2001; 2001WO-US15072.
 PF
 XX 10-MAY-2000; 2000US-203434P.
 PR
 XX (LOU) UNIV LOUISIANA STATE & AGRIC & MECH COLL.
 PA
 XX Croughan TP;
 PI
 XX WPI: 2002-097559/13.
 DR N-PSDB; ABK14657.
 DR
 XX Novel gene encoding a functional acetohydroxyacid synthase gene which
 PT imparts resistance to at least one herbicide, used for producing
 PT herbicide resistant rice -
 PS
 XX Example 29; Page 117-119; 157pp; English.
 XX
 XX This invention relates to novel nucleic acid and protein sequences of
 CC mutant acetohydroxyacid synthase (AHS) enzymes that can be used to
 CC create herbicide resistance green plants. The encoded AHS exhibits
 CC resistance to at least one herbicide as compared to the wild-type, and
 CC has a serine-asparagine substitution at amino acid 627. The sequences of
 CC the invention are useful creating herbicide resistance plants, by
 CC planting these plants it is easier to control the growth of weeds in the

CC vicinity of a plant by applying a herbicide which normally inhibits AHAS.
CC The sequences can be used for producing rice plants having resistance to
CC at least one herbicide which normally inhibits AHAS in the wild-type
CC plant, such herbicides may be particularly imidazolinone or
CC sulphonylurea herbicides. The production of plant having resistance to
CC AHAS inhibiting herbicide allows the development of new herbicides which
CC target AHAS, reducing the risk of weeds becoming resistant. The present
CC sequence represents the wild type protein sequence of the rice AHAS
CC enzyme used to create the herbicide resistant plants of the invention.
XX
SQ Sequence 644 AA;

Query Match 4.7%; Score 8; DB 23; Length 644;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
|||||
Db 6 AAAAAALS 13

RESULT 22

AAU10025
ID AAU10025 standard; Protein; 644 AA.

AC AAU10025;

DT 08-MAY-2002 (first entry)

DE Rice herbicide resistant AHAS protein sequence.

KW Rice; AHAS; acetohydroxyacid synthase; sulphonylurea herbicide;
KW herbicide resistance; weed control; imidazolinone; EC.4.1.3.18;
KW mutant; mutlein.

OS Oryza sativa.

PN WO200185970-A2.

PD 15-NOV-2001.

PF 09-MAY-2001; 2001WO-US15072.

PR 10-MAY-2000; 2000US-203434P.

PA (LOU) UNIV LOUISIANA STATE & AGRIC & MECH COLL.

PI Croughan TP;

DR WPI; 2002-097559/13.

DR N-PSDB; ABK14658.

PT Novel gene encoding a functional acetohydroxyacid synthase gene which
PT imparts resistance to at least one herbicide, used for producing
PT herbicide resistant rice.

XX Disclosure: Page 121-123; 157pp; English.

CC This invention relates to novel nucleic acid and protein sequences of
CC mutant acetohydroxyacid synthase (AHAS) enzymes that can be used to
CC create herbicide resistance green plants. The encoded AHAS exhibits
CC resistance to at least one herbicide as compared to the wild-type, and
CC has a serine-asparagine substitution at amino acid 627. The sequences of
CC the invention are useful creating herbicide resistance plants, by
CC planting these plants it is easier to control the growth of weeds in the
CC vicinity of a plant by applying a herbicide which normally inhibits AHAS.
CC The sequences can be used for producing rice plants having resistance to
CC at least one herbicide which normally inhibits AHAS in the wild-type
CC plant, such herbicides may be particularly imidazolinone or
CC sulphonylurea herbicides. The production of plant having resistance to
CC AHAS inhibiting herbicide allows the development of new herbicides which
CC target AHAS, reducing the risk of weeds becoming resistant. The present
CC sequence represents the sequence of the mutant protein sequence of the

CC rice AHAS enzyme used to create the herbicide resistant plants of the
CC invention.

SQ Sequence 644 AA;

Query Match 4.7%; Score 8; DB 23; Length 644;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
|||||
Db 6 AAAAAALS 13

RESULT 23

ABB59430
ID ABB59430 standard; Protein; 662 AA.

AC ABB59430;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 5082.

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL03533.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure: SEQ ID NO 5082; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 662 AA;

Query Match 4.7%; Score 8; DB 22; Length 662;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
|||||
Db 210 AAAAAALS 217

XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI: 2001-656860/75.
DR N-PSDB; ABL08785.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PI genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure: SEQ ID NO 20838; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 2090 AA:

Query Match 4.7%; Score 8; DB 22; Length 2090;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSL 15
DB 1349 AAAAALSL 1356

RESULT 27
ID ABB63299 standard; Protein; 2703 AA.
XX
AC ABB63299;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 16689.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR N-PSDB; ABL07402.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure: SEQ ID NO 16689; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB157737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 2703 AA:

Query Match 4.7%; Score 8; DB 22; Length 2703;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSL 15
DB 2647 AAAAALSL 2654

RESULT 28
ID AAW04637 standard; Peptide; 9 AA.
XX
AC AAW04637;
XX
DT 08-AUG-1997 (first entry)
XX
DE Null peptide that complexes with A2 protein.
XX
KW primer: PCR: polymerase chain reaction; specific inhibitor;
KW T-cell receptor; TCR: MHC; ligand; autoimmune disease; diabetes;
KW rheumatoid arthritis; Grave's disease; organ transplant rejection;
KW antigen; TCR-zeta; chimeric; CTL clone; HLA A2.1; HIV pol.
XX
OS Synthetic.
XX
PN WO9636881-A2.
XX
PD 21-NOV-1996.
XX
PF 16-MAY-1996; 96WO-GB01165.
XX
PR 16-MAY-1995; 95GB-0009844.
XX
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
XX
PI Barouch DH, Jakobsen BK, Vessey SJ;
XX
DR WPI: 1997-012213/01.
XX
PT Specific inhibitors of interaction between T cell receptor and MHC
PT peptide ligand - identified by incubating receptor expressing cells
PT with ligand and test cpd., and measuring change in interaction to
PT detect cpds. potentially useful for blocking disease related T cells
XX
PS Example -: Page 3; 21pp; English.
XX
CC Novel specific inhibitors of the interaction between a T-cell receptor
CC (TCR) and a MHC peptide ligand, are identified by: (a) incubating
CC responder cells that express TCR with a MHC ligand which stimulates the
CC cells, and a test cpd.; (b) monitoring a signal produced by the cell
CC when the TCR and MHC peptide ligand interact; and (c) comparing this
CC signal with a control signal from a similar system lacking the test
CC cpd.. The specific inhibitors are potentially useful for blocking
CC T-cells that cause autoimmune diseases (e.g. diabetes, rheumatoid
CC arthritis, Grave's disease etc.), organ transplant rejection or other T
CC cell mediated conditions. The interaction between a specific MHC peptide
CC ligand and a TCR was studied using purified MHC-single peptide complexes
CC (e.g. AAW04633-37) and an antigen specific TCR-zeta chimeric receptor
CC expressed on the surface of a basophil cell line. AAT43774-77 are primer

CC sequences used to amplify TCR genes by PCR from cDNA prepared from a
 CC CTL clone specific for HLA A2.1 restricted HIV pol peptide.
 XX
 SO Sequence 9 AA;

Query Match 4.1%; Score 7; DB 18; Length 9;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 |||||
 DB 3 AAAAAA 9

RESULT 29
 AAW61561
 ID AAW61561 standard; peptide: 10 AA.

XX AAW61561;

AC 19-OCT-1998 (first entry)

DE Biotinylated peptide GYR(A)6L.

XX T-cell; malaria; immunogenic; anti-malarial; prophylactic immunity.

OS Synthetic.

XX WO9831382-A1.

XX 23-JUL-1998.

XX 21-JAN-1998; 98WO-US01527.

XX 21-JAN-1997; 97US-0033916.

XX (UYNV) UNIV NEW YORK STATE.

XX Moreno A, Nardin E;

XX WPI: 1998-413810/35.

XX New immunogenic compositions for malaria - comprise malaria derived
 PT peptide comprising universal T-cell epitope which elicits
 PR anti-malarial T-cell response

XX Disclosure: Page 11; 38pp; English.

XX The biotinylated peptides AAW61559-W61561 were used in a peptide binding
 CC assay to test the peptide interactions of a T-cell epitope derived from
 CC malaria can be used in an immunogenic composition. The T-cell epitope
 CC elicits an anti-malarial T-cell response in mammals of diverse genetic
 CC backgrounds. The composition can be used as a vaccine to confer
 CC prophylactic or therapeutic immunity against malaria. They may also be
 CC used to inhibit the propagation of a malarial organism in a susceptible
 CC animal.

XX Sequence 10 AA;

Query Match 4.1%; Score 7; DB 19; Length 10;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 |||||
 DB 4 AAAAAA 10

RESULT 30
 AAR84046
 ID AAR84046 standard; peptide: 13 AA.
 XX
 AC AAR84046;

XX 14-MAY-1996 (first entry)

DE Murine MHC class II binding peptide E5:10.

XX Murine; MHC class II; binding peptide; haptenated peptides;
 KW contact; sensitivity; desensitising; mammal; allergen; ivy;
 KW urushiol; poison; oak.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "acylated"

FT Modified-site 13 /note= "amidated"

XX WO9526980-A2.

XX 12-OCT-1995.

XX 30-MAR-1995; 95WO-US04121.

XX 06-FEB-1995; 95US-0383645.

XX 01-APR-1994; 94US-0222206.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Gefter ML, Gelber C, Greenstein JL, Hackett CJ;

XX WPI: 1995-358583/46.

XX Haptenated peptide(s) capable of binding to Class II MHC molecules -
 PT for treating contact dermatitis
 CC Example: Fig 2; 85pp; English.

XX A peptide of 7-30 amino acids capable of binding to a murine MHC
 CC class II mol. (i.e. AAR84018-47) covalently linked to 1-3 hapten
 CC mols. can be used for treating contact sensitivity, or
 CC desensitising a mammal to a contact allergen (e.g. urushiol of
 CC poison ivy/oak). The peptide-hapten cpds. disrupt the normal
 CC proliferation of hapten-specific T cells, or alter the T cell
 CC mediated delayed-type hypersensitivity response to the hapten,
 CC resulting in effective desensitisation to the hapten.

XX Sequence 13 AA;

Query Match 4.1%; Score 7; DB 16; Length 13;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 IAAAAA 13
 |||||
 DB 3 IAAAAA 9

RESULT 31
 AAW02569
 ID AAW02569 standard; peptide: 14 AA.

XX AAW02569;

XX 13-JAN-1997 (first entry)

XX AKAP79 A37-50 mutation.

XX Autoimmune disease; PACT59; PACT74; PACT36; PACT60; murine; T-cell; PKA;
 KW mouse; A-kinase anchoring protein 79; cAMP-dependent protein kinase;
 KW postsynaptic density; AKAP79; human forebrain; transcriptional activator;
 KW interleukin 2; T-cell activation; calcineurin; T-cell clonal expansion;
 KW calcium/calmodulin dependent protein phosphatase; T-cell response;

KW autoimmune related disease; therapy; immune response.
 XX Synthetic.
 XX
 FT Key Location/Qualifiers
 FT Misc-difference 3
 FT Misc-difference 4 /note= "Arg39Ala"
 FT Misc-difference 5 /note= "Arg40Ala"
 FT Misc-difference 6 /note= "Lys41Ala"
 FT Misc-difference 9 /note= "Lys42Ala"
 FT Misc-difference 14 /note= "Lys45Ala"
 FT Misc-difference 14 /note= "Lys50Ala"
 XX
 PN W09616172-A2.
 XX
 PD 30-MAY-1996.
 XX
 PF 22-NOV-1995; 95MO-US16039.
 XX
 PR 17-JUL-1995; 95US-0503226.
 PR 23-NOV-1994; 94US-0344227.
 PR 15-MAR-1995; 95US-0404731.
 XX
 PA (ICOS-) ICOS CORP.
 PA (OREG-) STATE OF OREGON.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Coghlan VM, Gallatin WM, Howard ML, Lockerbie RO;
 PI Scott JD;
 XX
 DR WPI; 1996-268608/27.
 XX
 PT New modulators of anchoring protein function - used to develop
 PT prods. for use in the treatment of auto-immune-related conditions.
 XX
 PS Example 11; Page 34; 74pp; English.
 XX
 CC AAM02565-W02569 represent mutants of residues 37-50 of the A-kinase
 CC anchoring protein 79 (AKAP79). AKAP79 (see AAM05264 for wild type
 CC residues 37-50) is responsible for anchoring protein kinase
 CC (PKA) to specific intracellular sites. AKAP79 is predominantly present
 CC in postsynaptic densities in the human forebrain. The pathways that
 CC involve AKAP79 are important in many cell types and have been implicated
 CC in many cell functions, including the transcriptional activation of the
 CC interleukin 2 gene that is important in T-cell activation. AKAP also
 CC binds to calcineurin (see AAM02536), which is a calcium/calmodulin
 CC dependent protein phosphatase associated with T-cell activation. By
 CC binding both PKA and calcineurin, AKAP79 co-localises a kinase and a
 CC phosphatase which may regulate flux through a specific signalling
 CC pathway. The AKAP79 binding sequences can be used to develop products
 CC for use in the treatment of autoimmune related conditions. The AKAP79
 CC binding proteins can be used in methods for stimulating an immune
 CC response, and for stimulating activated T-cells for selected clonal
 CC expansion. The proteins can also be used in a method for enhancing
 CC T-cell responses to experimental stimuli for evaluation of early events
 CC in T-cell biology and activation of the immune response.
 CC
 SQ Sequence 14 AA;
 XX
 QY Query Match 4.1%; Score 7; DB 17; Length 14;
 DB Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 8 AAAAAA 14
 5 AAAAAA 11

RESULT 32
 AAB14909
 ID AAB14909 standard; Peptide; 14 AA.
 XX
 AC AAB14909;
 XX
 DT 08-JAN-2001 (first entry)
 XX
 DE Mutant peptide AKAP 79 A37-50.
 XX
 KW Human; calcineurin-binding peptide; A-kinase anchor protein; AKAP;
 KW AKAP 79; immunostimulant; interleukin 2 expression modulation;
 KW graft rejection; transplantation; T cell-mediated disorder; mutant.
 XX
 OS Homo sapiens..
 OS Synthetic.
 XX
 PN US6107104-A.
 XX
 PD 22-AUG-2000.
 XX
 PF 27-SEP-1996; 96US-0721458.
 XX
 PR 23-NOV-1994; 94US-0344227.
 PR 15-MAR-1995; 95US-0404731.
 PR 17-JUL-1995; 95US-0503226.
 XX
 PA (ICOS-) ICOS CORP.
 XX
 PI Lockerbie RO, Gallatin WM, Lai Y, Howard ML;
 XX
 DR WPI; 2000-578541/54.
 XX
 PT Novel calcineurin deletion mutant having calcineurin polypeptide
 PT sequence and binding A-kinase anchor proteins, for treating graft
 PT rejection following organ transplantation and T cell-mediated disorders
 XX
 PS Example 11; Column 21; 53pp; English.
 XX
 CC The present sequence is a mutant peptide derived from A-kinase anchor
 CC protein 79 (AKAP 79). It is expressed as a poly-histidine tag fusion
 CC protein and can thus be purified to homogeneity by nickel affinity
 CC chromatography. AKAP 79 binds both CAMP-dependent protein kinase (PKA)
 CC and calcineurin and so co-localises a kinase and a phosphatase that
 CC may regulate flux through a specific signalling pathway. Calcineurin is a
 CC Ca2+/calmodulin-dependent protein phosphatase which is involved in many
 CC intracellular signalling pathways. It participates in regulation of IL-2
 CC expression following T cell stimulation in T cells. Calcineurin-binding
 CC peptides derived from AKAP 79 may be used to inhibit calcineurin activity
 CC in a cell. The peptides are useful for treating graft rejection following
 CC organ transplantation and for treating T cell-mediated disorders.
 CC Calcineurin deletion mutants which bind AKAP 79 are useful for defining
 CC an AKAP 79 binding site, for stimulating the immune response, stimulating
 CC activated T cells for selected clonal expansion, or for enhancing T cell
 CC responses to experimental stimuli for evaluation of early events in
 CC T cell biology and activation of the immune response.
 CC
 SQ Sequence 14 AA;
 XX
 QY Query Match 4.1%; Score 7; DB 21; Length 14;
 DB Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 8 AAAAAA 14
 5 AAAAAA 11
 RESULT 33
 AAR84047
 ID AAR84047 standard; peptide; 16 AA.
 XX

```

AC AAR84047;
XX
XX 14-MAY-1996 (first entry)
XX
XX Murine MHC class II binding peptide E5:10:PDG.
DE
XX Murine MHC class II; binding peptide; haptenated peptides;
KM contact; sensitivity; desensitising; mammal; allergen; ivy;
KM urushiol; poison; oak.
XX
XX Synthetic.
XX
XX Key location/Qualifiers
XX Modified-site 1 /note= "acylated"
XX Modified-site 16 /note= "amidated"
XX
XX WO9526980-A2.
XX
XX 12-OCT-1995.
XX
XX 30-MAR-1995; 95WO-US04121.
XX
XX 06-FEB-1995; 95US-0383645.
XX 01-APR-1994; 94US-0222206.
XX
XX (IMMUNO-) IMMUNOLOGIC PHARM CORP.
XX
XX Gefter ML, Gelber C, Greenstein JL, Hackett CJ;
XX Wilson KJ;
XX
XX WPI. 1995-358583/46.
XX
XX Haptenated peptide(s) capable of binding to Class II MHC molecules -
XX for treating contact dermatitis
XX
XX Example; Fig 2; 85pp; English.
XX
XX A peptide of 7-30 amino acids capable of binding to a murine MHC
XX class II mol. (i.e. AAR84018-47) covalently linked to 1-3 hapten
XX mols. can be used for treating contact sensitivity, or
XX desensitising a mammal to a contact allergen (e.g. urushiol of
XX poison ivy/oak). The peptide-hapten cpds. disrupt the normal
XX proliferation of hapten-specific T cells, or alter the T cell
XX mediated delayed-type hypersensitivity response to the hapten,
XX resulting in effective desensitisation to the hapten.
XX
XX Sequence 16 AA:
SQ
XX
XX Query Match 4.1%; Score 7; DB 16; Length 16;
XX Best Local Similarity 100.0%; Pred. No. 17;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 IAAAAA 13
XX |||||
XX 3 IAAAAA 9
DB
XX
XX RESULT 34
XX ABB27942
XX ID ABB27942 standard; Peptide; 21 AA.
XX
XX ABB27942;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human peptide #593 encoded by breast cell single exon nucleic acid probe.
XX
XX Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer.
XX
XX Homo sapiens.
XX
XX OS

```

```

XX
XX WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00662.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 10910; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a peptide encoded by a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 21 AA:
SQ
XX
XX Query Match 4.1%; Score 7; DB 22; Length 21;
XX Best Local Similarity 100.0%; Pred. No. 22;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 130 NYLGGK 136
XX |||||
XX 4 NYLGGK 10
DB
XX
XX RESULT 35
XX ABB37770
XX ID ABB37770 standard; Peptide; 21 AA.
XX
XX ABB37770;
XX
XX 04-FEB-2002 (first entry)
XX
XX Peptide #5276 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
XX
XX WO200157271-A2.
XX
XX PN

```

XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00669.
 XX
 PE
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 27; SEQ ID NO 30405; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SQ Sequence 21 AA;
 SQ
 Query Match 4.1%; Score 7; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 NYLLGKG 136
 Db 4 NYLLGKG 10
 DE
 AC AAM53912 standard; Protein; 21 AA.
 AC AAM53912;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26017.
 XX
 KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX
 OS Homo sapiens.
 OS
 PN WO200157275-A2.
 PN
 PD 09-AUG-2001.
 PD
 PE 30-JAN-2001; 2001WO-US00667.
 PE
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 PS Example 4; SEQ ID NO: 26017; 650pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 CC
 SQ Sequence 21 AA;
 SQ
 Query Match 4.1%; Score 7; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 NYLLGKG 136
 Db 4 NYLLGKG 10
 DE
 AC AAM70864 standard; Protein; 21 AA.
 AC AAM70864;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31170.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 XX
 OS Homo sapiens.
 OS
 PN WO200157276-A2.
 PN
 PD 09-AUG-2001.
 PD
 PE 30-JAN-2001; 2001WO-US00668.
 PE
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 31170; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.

XX Sequence 21 AA;

Query Match 4.1%; Score 7; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 130 NYLKGK 136
Db 4 NYLKGK 10

RESULT 38
AAM18708
ID AAM18708 standard; Protein; 21 AA.

XX AC AAM18708;
XX DT 12-OCT-2001 (first entry)

DE Peptide #5142 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KM cervical cancer.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PL Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -

XX PS Claim 27; SEQ ID NO 23534; 487bp; English.

XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENP: see AAI10068-AI128459). The present sequence is a peptide encoded
XX CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
XX CC can be used to produce a single exon microarray, which can be used for
XX CC measuring human gene expression in a sample derived from human cervical
XX CC epithelial cells. By measuring gene expression, the probes are therefore
XX CC useful in grading and/or staging of diseases of the cervix, notably
XX CC cervical cancer.

XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SO Sequence 21 AA;

Query Match 4.1%; Score 7; DB 22; Length 21;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 130 NYLKGK 136
Db 4 NYLKGK 10

RESULT 39
AAM26577
ID AAM26577 standard; Protein; 21 AA.

XX AC AAM26577;

XX DT 17-OCT-2001 (first entry)

DE Peptide #614 encoded by probe for measuring placental gene expression.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KM genetic disorder.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-48897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -

XX PS Claim 27; SEQ ID NO 26846; 654bp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AAI31315-AI57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.

XX SO Sequence 21 AA;

Query Match 4.1%; Score 7; DB 22; Length 21;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 130 NYLKGK 136
Db 4 NYLKGK 10

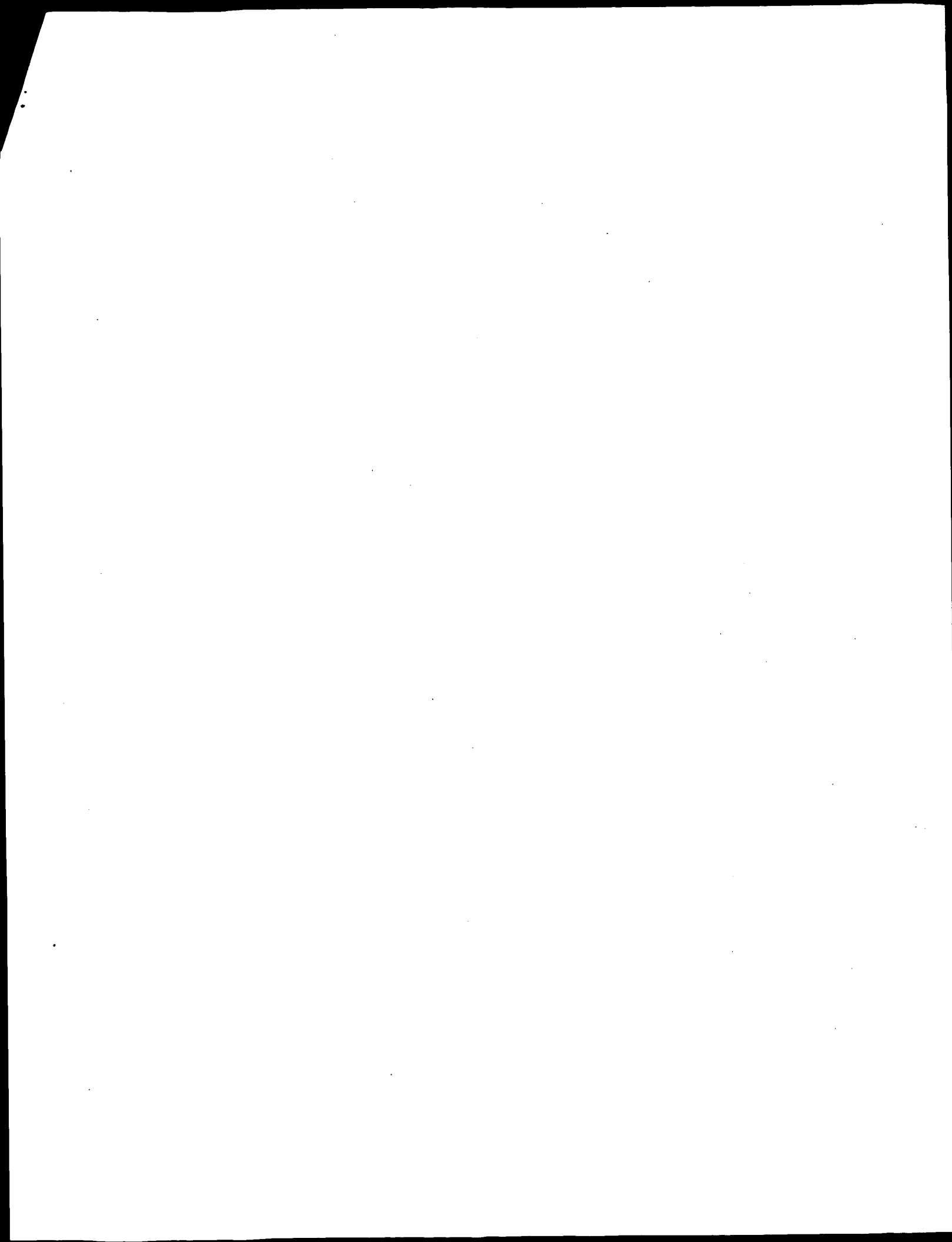
RESULT 40
ABG40662
ID ABG40662 standard; Peptide; 21 AA.

XX AC ABG40662;

XX DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 30327.
XX
KW Human: single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
XX 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2002-114183/15.
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 27; SEQ ID NO 30327; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression to a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 21 AA;
Query Match
Best Local Similarity 4.18; Score 7; DB 23; Length 21;
Matches 7; Conservative 100.0%; Pred. No. 22;
Mismatches 0; Indels 0; Gaps 0;
Db 130 NYYLGGK 136
4 NYYLGGK 10
Search completed: July 6, 2003, 14:20:17
Job time : 39 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

run on: July 6, 2003, 14:20:20 ; Search time 21 Seconds
(without alignments)
787.387 Million cell updates/sec

Title:	US-09-674-779B-2
Perfect score:	172

Sequence: 1 MMLHIQIAAAALSVLFM.....IAFGTNEEAWSQNRAELSY 172

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Maximum DB seq length: 20000000000

Post-processing: Listing first 100 summaries

```
Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	13	7.6	176	2	A60337	outer membrane protein
2	10	5.8	186	2	E82625	outer membrane protein
3	9	5.2	2176	2	T13806	toucan gene protein
4	8	4.7	143	2	H72296	sugar-phosphate isomerase
5	8	4.7	281	2	B95296	conserved hypothetical protein
6	8	4.7	380	2	JC7650	pectin lyase (EC 4.1.1.1)
7	8	4.7	452	2	H81777	UDP-N-acetylmutaromannose 4-epimerase
8	8	4.7	491	2	T40455	penicillin binding protein
9	8	4.7	526	2	G82981	probable binding protein
10	8	4.7	534	2	T50939	Dile protein (imp)
11	8	4.7	829	2	A34692	eodysone-induced protein
12	8	4.7	2715	2	T13049	eyelid - fruit fly
13	7	4.1	37	1	FDPL3W	antifreeze protein
14	7	4.1	40	1	FDF18G	antifreeze protein
15	7	4.1	82	1	FDF1AW	antifreeze protein
16	7	4.1	82	2	S02326	antifreeze protein
17	7	4.1	82	2	JS0706	antifreeze protein
18	7	4.1	82	2	AO5161	antifreeze protein
19	7	4.1	82	2	JS0705	antifreeze protein
20	7	4.1	82	2	IS1125	antifreeze protein
21	7	4.1	87	2	A55846	antifreeze protein
22	7	4.1	150	2	F86299	oppf homolog, kymdog
23	7	4.1	165	2	B70533	hypothetical protein
24	7	4.1	167	2	S58217	hypothetical protein
25	7	4.1	168	2	G83525	outer membrane protein
26	7	4.1	168	2	I40346	omp6 protein
27	7	4.1	168	2	AF3294	peptidoglycan-asso
28	7	4.1	170	2	AE2405	hypothetical protein
29	7	4.1	173	1	LPCEPG	peptidoglycan-asso

30	7	4.1	173	2	A85577	peptidoglycan-asso
31	7	4.1	173	2	A85725	peptidoglycan-asso
32	7	4.1	174	2	A10592	peptidoglycan-asso
33	7	4.1	174	2	F72550	hypothetical prote
34	7	4.1	176	2	T52661	cysteine synthase
35	7	4.1	177	2	AE83013	ompB protein [imp
36	7	4.1	177	2	B988271	ompB protein [imp
37	7	4.1	179	2	D75351	probable acetyltra
38	7	4.1	185	2	A62752	hypothetical prote
39	7	4.1	188	2	E71495	probable peptidogl
40	7	4.1	192	2	H22036	peptidoglycan-asso
41	7	4.1	192	2	C86588	peptidoglycan-asso
42	7	4.1	195	2	AG2727	conserved hypothet
43	7	4.1	200	2	G86214	protein T6D22.4 [i
44	7	4.1	202	2	C81653	peptidoglycan asso
45	7	4.1	203	2	I49153	cardiotrophin-1 -
46	7	4.1	213	2	A84250	NADH oxidase [imp
47	7	4.1	214	2	JC4808	ribosomal protein
48	7	4.1	214	2	E75613	hypothetical prote
49	7	4.1	233	2	D95860	probable transcrip
50	7	4.1	240	2	E72629	hypothetical prote
51	7	4.1	244	2	A98330	sporulation transcr
52	7	4.1	244	2	AD2953	transcription regu
53	7	4.1	251	2	B70765	probable cobM - My
54	7	4.1	260	2	C70675	hypothetical prote
55	7	4.1	260	2	E87310	pyrroline-5-carbox
56	7	4.1	261	2	S63604	homeobox protein G
57	7	4.1	267	2	B97509	hypothetical prote
58	7	4.1	267	2	S38367	tetrahydromethanop
59	7	4.1	267	2	G69021	tetrahydromethanop
60	7	4.1	271	2	H69097	phosphate-binding
61	7	4.1	272	2	AH2847	pyrroline-5-carbox
62	7	4.1	274	2	G97624	delta 1-pyrroline-
63	7	4.1	279	2	F84320	hypothetical prote
64	7	4.1	288	2	C81422	hypothetical prote
65	7	4.1	291	2	S05508	probable ATP-bind
66	7	4.1	294	2	H84115	photosystem II oxy
67	7	4.1	296	2	T30575	ribokinase rbsK [i
68	7	4.1	302	2	A12675	hypothetical prote
69	7	4.1	303	2	B87367	metallo-beta-lacta
70	7	4.1	303	2	F84401	ribokinase [impor
71	7	4.1	305	2	I57039	hypothetical prote
72	7	4.1	306	2	AB3431	genomic screen hom
73	7	4.1	308	2	D70875	hydroxyacylgutath
74	7	4.1	322	2	S23053	probable PE protei
75	7	4.1	322	2	S28125	sloppy paired prot
76	7	4.1	325	2	A37185	gas-vesicle operon
77	7	4.1	326	2	D72689	alpha-antigen prec
78	7	4.1	326	2	H97457	hypothetical prote
79	7	4.1	330	2	S26596	probable metallobe
80	7	4.1	331	2	E84770	Graves disease mit
81	7	4.1	333	2	A39065	probable serpin [i
82	7	4.1	335	2	S70671	homeotic protein E
83	7	4.1	335	2	B87590	lipopolysaccharide
84	7	4.1	336	2	T34783	hypothetical prote
85	7	4.1	339	2	T26328	probable signal pe
86	7	4.1	343	2	E64363	hypothetical prote
87	7	4.1	344	2	F70922	acidic ribosomal p
88	7	4.1	349	2	A44507	hypothetical prote
89	7	4.1	350	2	JC7188	lichenanase (EC 3.
90	7	4.1	353	2	B70358	lichenanase (EC 3.
91	7	4.1	357	2	JC6087	REIC protein - hum
92	7	4.1	370	2	E69754	hydrogenase (EC 1
93	7	4.1	374	2	B64775	helix-loop-helix t
94	7	4.1	388	2	C70605	nifs protein homol
95	7	4.1	389	2	C70605	hypothetical prote
96	7	4.1	394	2	T03411	probable aspartate
97	7	4.1	394	2	E75439	protein RFU - rice
98	7	4.1	401	2	H70605	conserved oxidore
99	7	4.1	403	2	A53662	hypothetical prote
100	7	4.1	404	2	T24255	homeotic protein H
						hypothetical prote

ALIGNMENTS

RESULT 1

A60337

outer membrane protein pPIA, peptidoglycan-associated, precursor - Legionella pneumophila

N:Alternate names: lipoprotein antigen

C:Species: Legionella pneumophila

C>Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 08-Oct-1999

C:Accession: A60337; S16531

R:Ludwig, B.; Schmid, A.; Marre, R.; Hacker, J.

A>Title: Cloning, genetic analysis, and nucleotide sequence of a determinant coding for

A:Reference number: A60337; MUID:91310296; PMID:1855972

A:Accession: A60337

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <LUD>

R:Engleberg, N.C.; Howe, D.C.; Rogers, J.E.; Arroyo, J.; Eisenstein, B.I.

Mol. Microbiol. 5, 2021-2029, 1991

A>Title: Characterization of a Legionella pneumophila gene encoding a lipoprotein antigen

A:Reference number: S16531; MUID:92114778; PMID:1766377

A:Accession: S16531

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <ENG>

A:Cross-references: EMBL:X60543; NID:944128; PIDN:CAA43033.1; PID:944129

C:Superfamily: outer membrane protein A

C:Keywords: membrane protein; surface antigen

Query Match

Best Local Similarity 7.6%; Score 13; DB 2; Length 176;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 AGHTDEGSRXN 117

DB 105 AGHTDEGSRXN 117

RESULT 2

E82625

outer membrane protein p6 precursor XPI896 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: E82625

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: E82625

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-186 <STM>

A:Cross-references: GB:AE004009; GB:AE003849; NID:91069980; PIDN:AA84702.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briano, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facinanci, A.P.; Ferreira, A.J.S.

Submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fromm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

C:Contents: annotation

A:Gene: XPI896

Query Match 5.8%; Score 10; DB 2; Length 186;
 Best Local Similarity 100.0%; Pred. No. 0.065;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 TDEGSRXN 117

DB 119 TDEGSRXN 128

RESULT 3

T13806

toucan gene protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13806

R:Grammont, M.; Dastugue, B.; Couderc, J.L.

Development 124, 4917-4926, 1997

A>Title: The Drosophila toucan (tcc) gene is required in germline cells for somatic c

A:Reference number: Z17769; MUID:98090047; PMID:9362455

A:Accession: T13806

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2176 <GBR>

A:Cross-references: EMBL:Y14157; NID:92760521; PIDN:CAA4574.1; PID:92760522

A:Gene: tcc

A:Cross-references: FlyBase:FBgn0015600

Query Match

Best Local Similarity 5.2%; Score 9; DB 2; Length 2176;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 16

DB 973 AAAAALSV 981

RESULT 4

H72296

sugar-phosphate isomerase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: H72296

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.

M. Nature 399, 323-329, 1999

A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: H72296

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-143 <ARN>

A:Cross-references: GB:AE001768; GB:AE000512; NID:94981619; PIDN:AAD36157.1; PID:9498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TMI080

C:Superfamily: galactoside O-acetyltransferase

Query Match

Best Local Similarity 4.7%; Score 8; DB 2; Length 143;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 NYLGGKI 137

DB 20 NYLGGKI 27

RESULT 5

B95296

conserved hypothetical protein SMA0520 [imported] - Sinorhizobium meliloti (strain 10

C:Species: Sinorhizobium meliloti

C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: B95296
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: B95296
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK4932.1; PID:914523354; GSPDB:GN00165
R:Gailbert, F.; Pihan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
P.; Chahin, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelure,
habault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma0520
A:Genome: plasmid

Query Match 4.7%; Score 8; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
DB 121 AAAAAAALS 128

RESULT 6

pectin lyase (EC 4.2.2.10) - *Aspergillus oryzae*
C:Species: *Aspergillus oryzae*
C:Date: 30-Jun-2001 #sequence, revision 30-Jun-2001 #text_change 02-Nov-2001
R:Kaimoto, N.; Yoshino-Yasuda, S.; Ohmiya, K.; Tsukagoshi, N.
Bioest. Biotechnol. Biochem. 65, 209-212, 2001
A:Title: Sequence analysis and overexpression of a pectin lyase gene (pell) from *Aspergillus*
A:Reference number: JC7650; MUID:21119720; PMID:11272833
A:Accession: JC7650
A:Molecule type: DNA
A:Residues: 1-381 <KIT>
A:Experimental source: strain KBN616
A:Accession: PC7125
A:Molecule type: protein
A:Residues: 21-35 <KIT>
C:Comment: This enzyme, a member of subclass of pectolytic enzymes, is the most important
ta-elimination, to form galacturonic oligomers with 4,5-unsaturated residues at the non-
C:Genetics:
A:Gene: pell
A:Introns: 159/22
C:Keywords: carbon-oxygen lyase

Query Match 4.7%; Score 8; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAAAAAL 14
DB 7 TAAAAAAL 14

RESULT 7

UDP-N-acetyluramylalanyl-D-glutamyl-2,6-diamino-pimelate-D-alanyl-D-alanine ligase (EC
C:Species: *Neisseria meningitidis*
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C:Accession: H81777

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: H81777
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <PAR>
A:Cross-references: GB:AL162758; GB:AL157959; NID:97380672; PIDN:CAB85286.1; PID:9738
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: murF; NMA2068
C:Keywords: ligase

Query Match 4.7%; Score 8; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
DB 289 AAAAAAALS 296

RESULT 8

penicillin binding protein pbp - *Bacillus subtilis*
C:Species: *Bacillus subtilis*
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
R:Rogoni, A.; Franchi, E.; Magistrelli, C.; Colombo, E.; Cosmina, P.; Grandi, G.
Microbiology 141, 645-648, 1995
A:Title: A putative new peptide synthase operon in *Bacillus subtilis*: partial charact
A:Reference number: I40454; MUID:95227362; PMID:7711903
A:Accession: I40455
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-491 <RES>
A:Cross-references: EMBL:Z34883; NID:91805667; PIDN:CAB4366.1; PID:9509467
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Ber
C:Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A:Ethlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
A:Authors: Fowler, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Leuber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, V.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: D69673
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-491 <KUN>
A:Cross-references: GB:Z99113; GB:AL009126; NID:92634090; PIDN:CAB13718.1; PID:926342
A:Experimental source: strain 168
C:Genetics:
A:Gene: pbp
C:Superfamily: D-alanyl-D-alanine carboxypeptidase

Query Match 4.7%; Score 8; DB 2; Length 491;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAAISVL 17
DB 88 AAAAISVL 95

RESULT 9

682981
Probable binding protein component of ABC dipeptide transporter PA5317 [Imported] - Pseu
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G82981
R:Stover, C.K.; Pham, X.O.; Errin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Llm,
; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G82981
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-526 <STO>
A:Cross-references: GB:AE004944; GB:AE004091; NID:99951628; PIDN:AA608702.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA5317
C:Superfamily: dipeptide transport protein

Query Match 4.7%; Score 8; DB 2; Length 526;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 AAAAALSV 16
|||||
Db 19 AAAAALSV 26

RESULT 10

T50939
DITE protein [Imported] - Pseudomonas abietaniphila
C:Species: Pseudomonas abietaniphila
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50939
R:Martin, V.J.; Mohu, W.W.
J: Bacteriol. 181, 2675-2682, 1999
A:Title: A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid-degrading
A:Reference number: Z25281; MUID:99235742; PMID:10217753
A:Accession: T50939
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-547 <MAR>
A:Cross-references: EMBL:AF119621; PIDN:AAD21067.1
A:Experimental source: strain BKME-9; ATCC700689
C:Genetics:
A:Gene: dITE

Query Match 4.7%; Score 8; DB 2; Length 547;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAALSVL 17
|||||
Db 107 AAAALSVL 114

RESULT 11

A34692
ecdysone-induced protein E74A - fruit fly (Drosophila melanogaster)
N:Alternate names: ets-related protein E74A
C:Species: Drosophila melanogaster
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 24-Sep-1998
C:Accession: A34692
R:Butt, K.C.; Thummel, C.S.; Jones, C.W.; Karim, F.D.; Hogness, D.S.
Cell 61, 85-99, 1990
A:Title: The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene
A:Reference number: A90912; MUID:90199900; PMID:2107982
A:Accession: A34692
A:Molecule type: mRNA

A:Residues: 1-829 <BUR>
A:Cross-references: GB:M37082; NID:9157307; PID:9157308
C:Genetics:
A:Gene: E74
A:Cross-references: FlyBase:FBgn0000567
C:Superfamily: ets DNA-binding domain homology
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation
F:735-815/Domain: ets DNA-binding domain homology <ETS>

Query Match 4.7%; Score 8; DB 2; Length 829;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAAAAL 14
|||||
Db 195 TAAAAAL 202

RESULT 12

T13049
eyelid - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13049
R:Freeman, J.E.; Luk, A.; Rubin, G.M.; Heberlein, U.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z17592
A:Accession: T13049
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2715 <TR>
A:Cross-references: EMBL:AF053091; NID:92981220; PID:92981221; PIDN:AAC06254.1
C:Genetics:
A:Gene: eld
A:Cross-references: FlyBase:FBgn0003013
C:Function:
A:Description: could act as a transcription factor antagonistic to the Wg pathway
C:Keywords: DNA binding

Query Match 4.7%; Score 8; DB 2; Length 2715;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALS 15
|||||
Db 2659 AAAAALS 2666

RESULT 13

FDPL3W
antifreeze protein 3 - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 23-Aug-1996
C:Accession: A03192
R:Devries, A.L.; Lin, Y.
Biochim. Biophys. Acta 495, 388-392, 1977
A:Title: Structure of a peptide antifreeze and mechanism of adsorption to ice.
A:Reference number: A03192; MUID:78060969; PMID:388591
A:Accession: A03192
A:Molecule type: protein
A:Residues: 1-37 <DEV>
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
|||||
Db 6 AAAAAL 12

RESULT 14

FDRI8G
antifreeze protein GS-8 - grubby sculpin
C:Species: Myoxocephalus aeneus (grubby sculpin)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 24-Nov-1999
C:Accession: S07046
R:Chakraborty, A.; Hew, C.L.; Shears, M.; Fletcher, G.
Can. J. Zool. 66, 403-408, 1988
A:Title: Primary structures of the alanine-rich antifreeze polypeptides from grubby sculpin
A:Reference number: S06417
A:Accession: S07046
A:Molecule type: protein
A:Residues: 1-40 <CH>
C:Superfamily: antifreeze protein
C:Keywords: antifreeze; blocked amino end
F:1/Modified site: blocked amino end (Met) #status experimental

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 14 AAAAAA 20

RESULT 15

FDPLAW

antifreeze protein A precursor - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 22-Jun-1999
C:Accession: JS0704; A03194
R:Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: JS0704
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:M62412; NID:g213592; PIDN:AAA49471.1; PID:g213593
R:Davies, P.L.; Roach, A.H.; Hew, C.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 335-339, 1982
A:Title: DNA sequence coding for an antifreeze protein precursor from winter flounder.
A:Reference number: A03194; MUID:82197490; PMID:6952188
A:Accession: A03194
A:Molecule type: mRNA
A:Residues: 1-82 <DA>
A:Experimental source: clones 4-2b and 2A-7c
A>Note: the authors translated the codon AGC for residue 24 as Arg
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze; plasma; tandem repeat
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-44/Domain: propeptide #status predicted <PRO>
F:45-82/Product: antifreeze protein A #status predicted <MAT>

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 50 AAAAAA 56

RESULT 16

S02326

antifreeze protein A - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 24-Oct-2000
C:Accession: S02326; JH0627
R:Scott, G.K.; Davies, P.L.; Kao, M.H.; Fletcher, G.L.

J. Mol. Evol. 27, 29-35, 1988

A:Title: Differential amplification of antifreeze protein genes in the Pleuronectinae
A:Reference number: S02326; MUID:88259236; PMID:3133486
A:Accession: S02326
A:Molecule type: DNA
A:Residues: 1-82 <SC>
A:Cross-references: EMBL:X07506; NID:964211; PIDN:CAA30389.1; PID:964212
R:Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: JH0627
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:M62415
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 50 AAAAAA 56

RESULT 17

JS0706

antifreeze protein (clone 4-2c) - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C>Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 24-Oct-2000
C:Accession: JS0706
R:Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: JS0706
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:M62417; NID:g213594; PIDN:AAA49472.1; PID:g213595
A>Note: the authors translated the codon AGC for residue 24 as Arg
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 50 AAAAAA 56

RESULT 18

A05161

antifreeze protein B precursor - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C>Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 24-Oct-2000
C:Accession: A05161
R:Davies, P.L.; Hough, C.; Scott, G.K.; Ng, N.; White, B.N.; Hew, C.L.
J. Biol. Chem. 259, 9241-9247, 1984
A:Reference number: A05161; MUID:84264559; PMID:6086629
A:Accession: A05161
A:Molecule type: DNA
A:Residues: 1-82 <DA>
A:Cross-references: GB:L00138; GB:J00929; NID:g343126; PIDN:AAB59964.1; PID:g457351
C:Genetics:
A:Introns: 19/2

C:Superfamily: antifreeze protein
C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 50 AAAAAAL 56

RESULT 19

JS0705

antifreeze protein (clones 1A-1a and 3-3a) - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)

C:Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 24-Oct-2000
C:Accession: JS0705

R: Davies, P.L.

Gene 112, 163-170, 1992

A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.

A:Reference number: JH0627; MUID:92209935; PMID:155765

A:Accession: JS0705

A:Molecule type: DNA

A:Residues: 1-82 <DAV>

A:Cross-references: GB:M62413; NID:g213586; PIDN:AAA49468.1; PID:g213587

C:Genetics:

A:Introns: 19/2

C:Superfamily: antifreeze protein

C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 50 AAAAAAL 56

RESULT 20

IS1125

antifreeze protein - winter flounder

C:Species: Pseudopleuronectes americanus (winter flounder)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Oct-2000

C:Accession: IS1125

R: Pickett, M., Scott, G., Davies, P., Wang, N., Joshi, S., Hew, C.

Eur. J. Biochem. 143, 35-38, 1984

A:Title: Sequence of an antifreeze protein precursor.

A:Reference number: IS1125; MUID:84285392; PMID:6547905

A:Accession: IS1125

A:Status: preliminary;

A:Molecule type: translated from GB/EMBL/DBJ

A:Residues: 1-82 <PIC>

A:Cross-references: GB:M8337; NID:g213581; PIDN:AAA49466.1; PID:g213582

C:Genetics:

A:Gene: AFP

C:Superfamily: antifreeze protein

C:Keywords: antifreeze

Query Match 4.1%; Score 7; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 50 AAAAAAL 56

RESULT 21

A55846

oprf homolog, kmb 5'-region - Streptomyces tenebrarius (fragment)
C:Species: Streptomyces tenebrarius

C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 05-May-1995
C:Accession: A55846

R: Holmes, D.J.; Cundliffe, E.

Mol. Gen. Genet. 229, 229-237, 1991

A:Title: Analysis of a ribosomal RNA methylase gene from Streptomyces tenebrarius whi

A:Reference number: S17717; MUID:9201656; PMID:1921972

A:Accession: A55846

A:Molecule type: DNA

A:Residues: 1-87 <HOL>

A:Note: The authors neither translated nor discussed this partial open reading frame

Query Match 4.1%; Score 7; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 RRAVAVR 129
|||||
Db 36 RRAVAVR 42

RESULT 22

F86299

hypothetical protein F309.23 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: F86299

R: Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marzla

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: F86299

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-150 <STO>

A:Cross-references: GB:AE005172; NID:g496364; PIDN:AA34695.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 4.1%; Score 7; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
|||||
Db 17 AAAAAAL 23

RESULT 23

B70533

hypothetical protein RW2719C - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: B70533

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

R: Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70533

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-165 <COL>

A:Cross-references: GB:Z96072; GB:AL123456; NID:g3261793; PIDN:CAB09460.1; PID:g21820

A:Experimental source: strain H37RV

C:Genetics:
A:Gene: RV2719c

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 165;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 VAPNAPT 39
DB 130 VAPNAPT 136

RESULT 24

outer membrane protein - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 26-Aug-1999

C:Accession: S58217
R:Lim, A.; de Vos, D.; Brauns, M.; Gaballa, A.; Hamers, R.; Cornelis, P.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
A:Reference number: S58216

A:Accession: S58217
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-167 <Lim>
A:Cross-references: EMBL:250191
C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 167;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 GHTDERG 112
DB 102 GHTDERG 108

RESULT 25

outer membrane protein OprL precursor PA0973 [imported] - Pseudomonas aeruginosa (strain C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G83525
R:Stover, C.K.; Pham, X.Q.; Ewitt, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; Lim, Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: G83525
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-168 <STO>
A:Cross-references: GB:AE004530; GB:AE004091; NID:g9946865; PIDN:AA04362.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:

A:Gene: oprL, PA0973
C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 168;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 GHTDERG 112
DB 102 GHTDERG 108

RESULT 26

ompL6 protein - Brucella abortus
C:Species: Brucella abortus

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40346

R:Libor, A.; Weynants, V.; Denoel, P.; Lichtfouse, B.; De Bolle, X.; Saman, E.; Limet Infect. Immun. 62, 3633-3639, 1994

A:Title: Molecular cloning, nucleotide sequence, and occurrence of a 16.5-kilodalton
A:Reference number: I40346; MUID:94341663; PMID:8063379

A:Accession: I40346
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

A:Residues: 1-168 <RES>
A:Cross-references: GB:L27996; NID:g619644; PIDN:AA59360.1; PID:g538293
C:Genetics:

A:Gene: pal
C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 168;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 WSONRRA 168
DB 153 WSONRRA 159

RESULT 27

peptidoglycan-associated lipoprotein [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Mar-2002

C:Accession: AF3294
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688

A:Accession: AF3294
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-168 <KUN>
A:Cross-references: GB:AE008917; PIDN:AA151521.1; PID:g17982237; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:

A:Gene: BMEI0340
A:Map position: I
C:Superfamily: outer membrane protein A

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 168;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 WSONRRA 168
DB 153 WSONRRA 159

RESULT 28

hypothetical protein al14797 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AE2405
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigun Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE2405
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-170 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA076496.1; PID:g17133934; GSPDB:GN00179
A:Experimental source: strain PCC 7120

C:Genetics:
A:Gene: all4797

Query Match 4.1%; Score 7; DB 2; Length 170;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 TGAVAPLV 53
|||||
Db 58 TGAVAPLV 64

RESULT 29

peptidoglycan-associated lipoprotein precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli
C:Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 01-Mar-2002
C:Accession: A27534; S20547; D64810
R:Chen, R.; Henning, U.
Eur. J. Biochem. 163, 73-77, 1987

A:Title: Nucleotide sequence of the gene for the peptidoglycan-associated lipoprotein of
A:Reference number: A27534; MUID:8713576; PMID:3545827
A:Accession: A27534

A:Molecule type: DNA
A:Residues: 1-173 <CH>

A:Cross-references: GB:X05123; NID:g42256; PIDN:CAA28771.1; PID:g42257
R:Laazaroni, J.C.; Portaiter, R.
Mol. Microbiol. 6, 735-742, 1992

A:Title: The exc gene of Escherichia coli K-12 required for cell envelope integrity and
A:Reference number: S20546; MUID:92244043; PMID:1574003
A:Accession: S20547

A:Molecule type: DNA
A:Residues: 1-173 <LMT>

A:Cross-references: EMBL:X65796; NID:g41358; PIDN:CAA6673.1; PID:g41360
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-
A.; Rose, D.J.; Mau, B.; Shaoy, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64810

A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-173 <BLAT>

A:Cross-references: GB:AB000177; GB:U00096; NID:g1786955; PIDN:AACT3835.1; PID:g1786962;
A:Experimental source: strain K-12, substrain MG1655
C:Comment: This lipoprotein of unknown function is very strongly associated with the per-
C:Genetics:

A:Gene: pal; excC
A:Map position: 17 min
C:Superfamily: outer membrane protein A

C:Keywords: lipid binding; lipoprotein
F.1-21/Domain: signal sequence #status predicted <Sig>
F.22-173/Product: peptidoglycan-associated lipoprotein #status predicted <Mat>

Query Match 4.1%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 SLGERRA 125
|||||
Db 120 SLGERRA 126

RESULT 30

peptidoglycan-associated lipoprotein [imported] - Escherichia coli (strain 0157:H7, subs-
A85577

C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 01-Mar-2002
C:Accession: A85577

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: A85577

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-173 <STO>

A:Cross-references: GB:AE005174; NID:g12513674; PIDN:ANG55077.1; GSPDB:GN00145; UWGP-
A:Experimental source: strain 0157:H7, substrain EDL933

C:Genetics:

A:Gene: pal
C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 SLGERRA 125
|||||
Db 120 SLGERRA 126

RESULT 31

peptidoglycan-associated lipoprotein [imported] - Escherichia coli (strain 0157:H7, s-
H90725

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 01-Mar-2002
C:Accession: H90725

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H90725

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-173 <LMT>

A:Cross-references: GB:BA000007; PIDN:BA84199.1; PID:g13360235; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecs0776

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 SLGERRA 125
|||||
Db 120 SLGERRA 126

RESULT 32

peptidoglycan-associated lipoprotein precursor [imported] - Salmonella enterica subsp
A10592

C:Species: Salmonella enterica subsp. enterica serovar Typh
A:Note: this species has also been called Salmonella typh
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 01-Mar-2002

C:Accession: A10592
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, K.M.; Dowd, L.; White, N.; Farr
S.; Moulé, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: A10592

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-174 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05211.1; PID:g16501981; GSPDB:GN00176
C:Genetics:
A:Gene: STY0795
C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SLGERA 125
|||||
Db 121 SLGERA 127

RESULT 33

F72550

hypothetical protein APE1690 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: F72550

R:Kawabadyai, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: F72550

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-174 <KAM>

A:Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAAB0691.1; PID:g5105378

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1690

C:Superfamily: Aeropyrum pernix hypothetical protein APE1690

Query Match 4.1%; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAAAAA 13
|||||
Db 89 TAAAAA 95

RESULT 34

T52661

cysteine synthase (EC 4.2.99.8), mitochondrial [imported] - Arabidopsis thaliana (fragme

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000

C:Accession: T52661

R:Hatzfeld, Y.; Maruyama, A.; Schmidt, A.; Noji, M.; Ishizawa, K.; Saito, K.

A>Title: Beta-cyanoalanine synthase is a mitochondrial cysteine synthase-like protein in

A:Reference number: 226167

A:Accession: T52661

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-176 <HAT>

A:Cross-references: EMBL:AF011044; PIDN:CAB55622.1

A:Experimental source: cultivar Columbia

C:Genetics:

A:Gene: oas6

C:Superfamily: threonine dehydratase

C:Keywords: carbon-oxygen lyase; mitochondrion

Query Match 4.1%; Score 7; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TAAAAA 14
|||||
Db 127 TAAAAA 133

RESULT 35

AE3013

omp16 protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Mar-2002

C:Accession: AE3013

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AE3013

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-177 <KUR>

A:Cross-references: GB:AE008689; PIDN:AA144523.1; PID:g17742135; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: palA

A:Map position: linear chromosome

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 WSONRRA 168
|||||
Db 162 WSONRRA 168

RESULT 36

B98271

omp16 protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)

C:Species: Agrobacterium tumefaciens

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 01-Mar-2002

C:Accession: B98271

R:Goodner, B.; Hinkle, G.; Gatung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm

A.; Liu, F.; Mollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001

A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: B98271

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-177 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK89692.1; PID:g15159600; GSPDB:GN00170

C:Genetics:

A:Gene: AGK_L_2246

A:Map position: linear chromosome

C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 WSONRRA 168
|||||
Db 162 WSONRRA 168

RESULT 37

D75351

probable acetyltransferase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: D75351

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: D75351

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <NH2>
A:Cross-references: GB:AE002021; GB:AE000513; NID:g6459573; PIDN:AAF11356.1; PID:g645957
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1800
A:Map position: 1
C:Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase rimJ

Query Match 4.1%; Score 7; DB 2; Length 179;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AAAAAA 15
|||||
DB 108 AAAAAA 114

RESULT 38

hypothetical protein XF0883 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: AB2752
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below

A:Accession: AB2752
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-185 <STM>
A:Cross-references: GB:AE003927; GB:AE003849; NID:g9105783; PIDN:AAF83693.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brlones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H
as-Neto, E.; Docena, C.; El-Doroty, H.; Faciniani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; From
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.U. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0883

Query Match 4.1%; Score 7; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
|||||
DB 93 AAAAAA 99

RESULT 39

probable peptidoglycan-associated lipoprotein - Chlamydia trachomatis (serotype D, strai
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 17-Mar-2000
C:Accession: E71495
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A:Reference number: AV1570; MUID:9500809; PMID:9784136
A:Accession: E71495

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <ARN>
A:Cross-references: GB:AE001330; GB:AE001273; NID:g3329034; PIDN:AA68202.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pal
C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 188;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 GHTDERG 112
|||||
DB 119 GHTDERG 125

RESULT 40

peptidoglycan-associated lipoprotein CP1091 [imported] - Chlamydia pneumoniae (st
H72036
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 19-May-2000
C:Accession: H72036; C81504
R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: H72036
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <ARN>
A:Cross-references: GB:AE001659; GB:AE001363; NID:g4377081; PIDN:AA018919.1; PID:g437
A:Experimental source: strain CWD029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gynn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salze
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: C81504
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <RNA>
A:Cross-references: GB:AE002265; GB:AE002161; NID:g7189999; PIDN:AAF38862.1; PID:g719
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pal; CP1091
C:Superfamily: outer membrane protein A

Query Match 4.1%; Score 7; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 GHTDERG 112
|||||
DB 123 GHTDERG 129

Search completed: July 6, 2003, 14:22:04
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 13:11:10 ; Search time 61 Seconds

(without alignments)
375.723 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 861
Sequence: 1 MMHIOIAAAALSLVTFM.....IARGTNEAMSONRAELSY 172

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database :

1: /SID52/gcgdata/geneseq/emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	861	100.0	172	21	AAV55089
2	855	99.3	172	21	AAV55090
3	850	98.7	172	21	AAV55091
4	844	98.0	172	21	AAV55092
5	266	30.9	153	11	AAV55092
6	264	30.7	153	9	AAV55093
7	264	30.7	153	9	AAV55094
8	264	30.7	153	11	AAV55095
9	264	30.7	153	15	AAV55096
10	259	30.1	153	9	AAV55097

11	215.5	25.0	179	18	AAW20394	H. pylori outer me
12	215.5	25.0	179	18	AAW24651	H. pylori outer me
13	215.5	25.0	179	20	AAW89962	Expressed antigen
14	215.5	25.0	179	22	AAW63316	H. pylori HPS144 p
15	215.5	25.0	179	22	AAW20795	H. pylori outer me
16	211.5	24.6	179	20	AAW89813	Protein encoded by
17	208.5	23.9	179	18	AAW23592	H. pylori omp22 ou
18	205.5	23.8	179	14	AAW23591	H. pylori recombin
19	204.5	23.8	179	14	AAW23591	Chlamydia pneumoni
20	204.5	23.8	179	14	AAW23591	Chlamydia pneumoni
21	204.5	23.8	179	14	AAW23591	Expressed antigen
22	197.5	22.9	179	23	AAW90544	Chlamydia trachoma
23	196.5	22.8	179	23	AAW90544	Moraxella catarrha
24	180.5	21.0	179	20	AAW90544	MOMP p5. Haemophil
25	176.5	20.5	179	20	AAW90544	P. aeruginosa orpf
26	175.5	20.4	179	20	AAW90544	Nonlymph H. infl
27	174.5	20.3	179	20	AAW90544	Protein associated
28	174.5	20.2	179	20	AAW90544	P. aeruginosa orpf
29	171.5	19.9	179	20	AAW90544	Outer membrane pro
30	171.5	19.9	179	20	AAW90544	Protein associated
31	171.5	19.9	179	20	AAW90544	Actinobacillus ple
32	171.5	19.9	179	20	AAW90544	Actinobacillus ple
33	170.5	19.7	179	20	AAW90544	Actinobacillus ple
34	170.5	19.6	179	20	AAW90544	Actinobacillus ple
35	169.5	19.6	179	20	AAW90544	Actinobacillus ple
36	165.5	19.2	179	20	AAW90544	Actinobacillus ple
37	165.5	19.2	179	20	AAW90544	Actinobacillus ple
38	161.5	18.8	179	20	AAW90544	Actinobacillus ple
39	161.5	18.8	179	20	AAW90544	Actinobacillus ple
40	161.5	18.8	179	20	AAW90544	Actinobacillus ple
41	161.5	18.8	179	20	AAW90544	Actinobacillus ple
42	160.5	18.6	179	20	AAW90544	Actinobacillus ple
43	158.5	18.4	179	20	AAW90544	Actinobacillus ple
44	158.5	18.4	179	20	AAW90544	Actinobacillus ple
45	158.5	18.4	179	20	AAW90544	Actinobacillus ple
46	158.5	18.4	179	20	AAW90544	Actinobacillus ple
47	155.5	18.1	179	20	AAW90544	Actinobacillus ple
48	155.5	18.1	179	20	AAW90544	Actinobacillus ple
49	155.5	18.1	179	20	AAW90544	Actinobacillus ple
50	150.5	17.5	179	20	AAW90544	Actinobacillus ple
51	147.5	17.1	179	20	AAW90544	Actinobacillus ple
52	145.5	16.8	179	20	AAW90544	Actinobacillus ple
53	145.5	16.8	179	20	AAW90544	Actinobacillus ple
54	144.5	16.7	179	20	AAW90544	Actinobacillus ple
55	144.5	16.7	179	20	AAW90544	Actinobacillus ple
56	142.5	16.6	179	20	AAW90544	Actinobacillus ple
57	142.5	16.6	179	20	AAW90544	Actinobacillus ple
58	142.5	16.6	179	20	AAW90544	Actinobacillus ple
59	138.5	16.1	179	20	AAW90544	Actinobacillus ple
60	138.5	16.1	179	20	AAW90544	Actinobacillus ple
61	136.5	15.9	179	20	AAW90544	Actinobacillus ple
62	136.5	15.9	179	20	AAW90544	Actinobacillus ple
63	136.5	15.9	179	20	AAW90544	Actinobacillus ple
64	136.5	15.9	179	20	AAW90544	Actinobacillus ple
65	136.5	15.9	179	20	AAW90544	Actinobacillus ple
66	136.5	15.9	179	20	AAW90544	Actinobacillus ple
67	136.5	15.9	179	20	AAW90544	Actinobacillus ple
68	136.5	15.9	179	20	AAW90544	Actinobacillus ple
69	136.5	15.9	179	20	AAW90544	Actinobacillus ple
70	136.5	15.9	179	20	AAW90544	Actinobacillus ple
71	136.5	15.9	179	20	AAW90544	Actinobacillus ple
72	136.5	15.9	179	20	AAW90544	Actinobacillus ple
73	136.5	15.9	179	20	AAW90544	Actinobacillus ple
74	136.5	15.9	179	20	AAW90544	Actinobacillus ple
75	136.5	15.9	179	20	AAW90544	Actinobacillus ple
76	136.5	15.9	179	20	AAW90544	Actinobacillus ple
77	136.5	15.9	179	20	AAW90544	Actinobacillus ple
78	136.5	15.9	179	20	AAW90544	Actinobacillus ple
79	136.5	15.9	179	20	AAW90544	Actinobacillus ple
80	136.5	15.9	179	20	AAW90544	Actinobacillus ple
81	136.5	15.9	179	20	AAW90544	Actinobacillus ple
82	130.5	15.2	179	20	AAW90544	Actinobacillus ple
83	122.5	14.2	179	20	AAW90544	Actinobacillus ple

84	122.5	14.2	550	22	ABB52693	Escherichia coli p
85	107.5	12.5	390	23	AAU09397	Soluble Porphyrom
86	107.5	12.5	391	20	AAV34490	Porphyromonas ging
87	107.5	12.5	395	20	AAV34365	Porphyromonas ging
88	107	12.4	380	23	AAU09398	Soluble Porphyrom
89	107	12.4	385	20	AAV34491	Porphyromonas ging
90	107	12.4	387	20	AAV34366	Porphyromonas ging
91	99.5	11.6	196	22	AAU03573	P. gingivalis chim
92	99.5	11.6	315	22	ABG17777	Novel human diagno
93	96	11.1	368	20	AAV19945	B. burgdorferi ant
94	96	11.1	389	20	AAV19944	Hepatitis GB virus
95	87.5	10.2	507	16	AAAB82451	Hepatitis GB virus
96	87.5	10.2	507	21	AAAB09475	Hepatitis GB virus
97	87.5	10.2	1422	16	AAAB82068	Hepatitis GB virus
98	87.5	10.2	1422	21	AAAB09038	Hepatitis GB virus
99	87.5	10.2	2864	16	AAAB82072	Hepatitis GB virus
100	87.5	10.2	2864	21	AAAB09268	Hepatitis GB virus

ALIGNMENTS

RESULT 1
AAV55089 standard; Protein: 172 AA.

AAV55089;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #1.

BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
genetic mutation screening; antibody production; vaccine; otitis media;
bacterial infection; pneumonia; sinusitis; nosocomial infection;
invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

MO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMRK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AA240351.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 101pp: English.

This sequence is a Moraxella catarrhalis BASB019 protein of the
invention. The sequences can be used for diagnosis of disease, staging of
disease, or determining response of an infectious organism to drugs. The
polynucleotides may be used as a source for hybridisation probes, and for
screening of genetic mutations, serotype, organism or strain
identification. Identification of mutation in BASB013 sequences, and as
components of arrays which are useful for diagnostic and prognostic
purposes. The polypeptides can be used to produce antibodies, and as a
target for the screening of antimicrobial drugs. The polypeptides can
also be used in vaccine formulations, and to identify agonists and
antagonists. The polypeptides, antibodies, agonists and antagonists
(which are bacteriostatic) are used for the treatment and prevention of
diseases including bacterial infection, otitis media in infants and

CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases; chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The frequency of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

Sequence 172 AA;
Query Match 100.0%; Score 861; DB 21; Length 172;
Best Local Similarity 100.0%; Pred. No. 87e-86;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLHIOIAAAALSVTFMTGCANKSTQVAVAPPGYVYTGAVPLVNDENVK 60
DB 1 MMLHIOIAAAALSVTFMTGCANKSTQVAVAPPGYVYTGAVPLVNDENVK 60
QY 61 ALASKLPSTLYVDFEDSEDEIKPQAAALDEQAQFLTTNOTARVLAGHTDERGSEYKNSL 120
DB 61 ALASKLPSTLYVDFEDSEDEIKPQAAALDEQAQFLTTNOTARVLAGHTDERGSEYKNSL 120
QY 121 GERRAVAVRNYLLGKGINQASVELISFGEGERPIAGTNEANSQRRELISY 172
DB 121 GERRAVAVRNYLLGKGINQASVELISFGEGERPIAGTNEANSQRRELISY 172

RESULT 2

AAV55090 standard; Protein: 172 AA.

AAV55090;

01-MAR-2000 (first entry)

M. catarrhalis BASB019 protein sequence #2.

BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
genetic mutation screening; antibody production; vaccine; otitis media;
bacterial infection; pneumonia; sinusitis; nosocomial infection;
invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

Moraxella catarrhalis.

MO9957277-A2.

11-NOV-1999.

03-MAY-1999; 99WO-EP03038.

06-MAY-1998; 98GB-0009683.

(SMRK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI: 2000-062148/05.

N-PSDB: AA240352.

Novel BASB019 polynucleotides and polypeptides from Moraxella

catarrhalis used to prepare vaccines against bacterial infections

Claim 3; Fig 3; 101pp: English.

This sequence is a Moraxella catarrhalis BASB019 protein of the

invention. The sequences can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutation in BASB019 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies, and as a target for the screening of antimicrobial drugs. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases including bacterial infection, otitis media in infants and children, pneumonia in the elderly, sinusitis, nosocomial infections and invasive diseases, chronic otitis media with hearing loss, fluid accumulation in the middle ear, auditive nerve damage, delayed speech learning, infection of the upper respiratory tract and middle ear infection. They are also used in the prevention of adhesion of bacteria to eukaryotic matrix proteins on in-dwelling devices, or to extracellular proteins on wounds, and to thus prevent tissue damage and/or block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques, and it is no longer common to isolate *M. catarrhalis* strains that are resistant to standard antibiotics. The BASB019 products of the invention can be used screen for new antibacterial compounds that may target these resistant bacteria.

Sequence 172 AA;

Query Match 99.3%; Score 85; DB 21; Length 172;
Best Local Similarity 99.4%; Pred. No. 4e-85;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MMHIOIAAAALSVLTFMTGCANKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60
Db 1 MMHIOIAAAALSVLTFMTGCANKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60
QY 61 ALASKPLSLVYDFPDSDEIKPOAAALIDEOAQFLTTNOTARVLAAGHTDGRSGREYNMSL 120
Db 61 ALASTPLSLVYDFPDSDEIKPOAAALIDEOAQFLTTNOTARVLAAGHTDGRSGREYNMSL 120
QY 121 GERAAVAVRNLYLKGINGQASVEIISFGERPFIAGTNEAMSONRRALSTY 172
Db 121 GERAAVAVRNLYLKGINGQASVEIISFGERPFIAGTNEAMSONRRALSTY 172

RESULT 3

AAVS5091
ID AAVS5091 standard; Protein; 172 AA.

AC AAVS5091;
DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein sequence #3.

KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

XX Moraxella catarrhalis.

OS W09957277-A2.

PN 11-NOV-1999.

PD 03-MAY-1999; 99WO-EP03038.

PF 06-MAY-1998; 98GB-0009683.

PR (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;
PI WPI: 2000-062148/05.
DR N-PSDB: AAZ40353.

PT Novel BASB019 polynucleotides and polypeptides from Moraxella
catarrhalis used to prepare vaccines against bacterial infections
PS Claim 3; Fig 3; 101pp; English.

This sequence is a Moraxella catarrhalis BASB019 protein of the invention. The sequences can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutation in BASB019 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies, and as a target for the screening of antimicrobial drugs. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases including bacterial infection, otitis media in infants and children, pneumonia in the elderly, sinusitis, nosocomial infections and invasive diseases, chronic otitis media with hearing loss, fluid accumulation in the middle ear, auditive nerve damage, delayed speech learning, infection of the upper respiratory tract and middle ear infection. They are also used in the prevention of adhesion of bacteria to eukaryotic matrix proteins on in-dwelling devices, or to extracellular proteins on wounds, and to thus prevent tissue damage and/or block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques, and it is no longer common to isolate *M. catarrhalis* strains that are resistant to standard antibiotics. The BASB019 products of the invention can be used screen for new antibacterial compounds that may target these resistant bacteria.

Sequence 172 AA;

Query Match 98.7%; Score 85; DB 21; Length 172;
Best Local Similarity 98.3%; Pred. No. 1.4e-84;
Matches 169; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MMHIOIAAAALSVLTFMTGCANKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60
Db 1 MMHIOIAAAALSVLTFMTGCANKSTSQVAVAPNPTGYTYGVAPLVNDDETVK 60

QY 61 ALASKPLSLVYDFPDSDEIKPOAAALIDEOAQFLTTNOTARVLAAGHTDGRSGREYNMSL 120
Db 61 TLASTPLSLVYDFPDSDEIKPOAAALIDEOAQFLTTNOTARVLAAGHTDGRSGREYNMSL 120

QY 121 GERAAVAVRNLYLKGINGQASVEIISFGERPFIAGTNEAMSONRRALSTY 172
Db 121 GERAAVAVRNLYLKGINGQASVEIISFGERPFIAGTNEAMSONRRALSTY 172

RESULT 4

AAVS5092
ID AAVS5092 standard; Protein; 172 AA.

AC AAVS5092;

DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein sequence #4.

KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
upper respiratory tract infection; middle ear infection; therapy.

XX Moraxella catarrhalis.
 OS W09957277-A2.
 PN 11-NOV-1999.
 PD 03-MAY-1999; 99WO-EP03038.
 PE 06-MAY-1998; 98GB-0009683.
 PR (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Ruelle J;
 XX WPI: 2000-062148/05.
 DR N-PSDB; AA240354.
 XX
 PT Novel BASB019 polynucleotides and polypeptides from Moraxella
 catarrhalis used to prepare vaccines against bacterial infections
 PS Claim 3; Fig 3; 101pp; English.

CC This sequence is a Moraxella catarrhalis BASB019 protein of the
 CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a
 CC target for the screening of antimicrobial drugs. The polypeptides can
 CC also be used in vaccine formulations, and to identify agonists and
 CC antagonists. The polypeptides, antibodies, agonists and antagonists
 CC (which are bacteriostatic) are used for the treatment and prevention of
 CC diseases including bacterial infection, otitis media in infants and
 CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
 CC invasive diseases, chronic otitis media with hearing loss, fluid
 CC accumulation in the middle ear, auditive nerve damage, delayed speech
 CC learning, infection of the upper respiratory tract and middle ear
 CC infection. They are also used in the prevention of adhesion of bacteria
 CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
 CC proteins on wounds, and to thus prevent tissue damage and/or block the
 CC normal progression of pathogenesis in infections initiated other than by
 CC the implantation of in-dwelling devices or by other surgical techniques.
 CC The frequency of Moraxella catarrhalis infections has risen dramatically,
 CC and it is no longer common to isolate M. catarrhalis strains that are
 CC resistant to standard antibiotics. The BASB019 products of the invention
 CC can be used to screen for new antibacterial compounds that may target these
 CC resistant bacteria.

XX Sequence 172 AA;

Query Match 98.0%; Score 844; DB 21; Length 172;

Best Local Similarity 98.3%; Pred. No. 6, 3e-84;
 Matches 169; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MMLHIQIAAAAAALSVLFTMGCAKSTQVAVANAPFGYGVLYTGVAPLVNDETVK 60
 DB 1 MMLHIQIAAAAAALSVLFTMGCAKSTQVAVANAPFGYGVLYTGVAPLVNDETVK 60
 QY 61 ALASKLPSLVYFDFSDSEIKPQAAALIDEOAFLTTNQTARVLVAGHTDGRSGREYNM 120
 DB 61 ALASTLPSLVYFDFSDSEIKPQAAALIDEOAFLTTNQTARVLVAGHTDGRSGREYNM 120
 QY 121 GERRAVAVRNYLLGKGINQASVETISFGEERPIAGTNEASQRRRAELSY 172
 DB 121 GERRAVAVRNYLLGKGINQASVETISFGEERPIAGTNEASQRRRAELSY 172

RESULT 5
 AAR07145
 ID AAR07145 standard; protein; 153 AA.

XX AAR07145;

AC 24-JAN-1991 (first entry)

DE 16.6KD outer membrane protein (OMP) of H. influenzae.

XX HI vaccine; OMP.

OS Haemophilus influenzae.

XX EP389925-A.

PD 03-OCT-1990.

PF 20-MAR-1990; 90EP-0105205.

XX 29-MAR-1989; 89US-0330229.

XX (UYNE-) STATE UNIV NEW YORK.

PA Murphy TF, Apicella MA;

XX WPI: 1990-298924/40.

DR N-PSDB; AAR07145.

XX Purification of outer membrane protein of haemophilus influenzae
 PT - by separation from an insoluble fraction using
 PT detergent-contg., then detergent-free buffers.

PS Disclosure; Page 8; 22pp; English.

CC Method claimed produces large quantities of the purified OMP, useful
 CC in raising antibodies for detection, and as a vaccine against
 CC H. influenzae.

XX Sequence 153 AA;

Query Match 30.9%; Score 266; DB 11; Length 153;

Best Local Similarity 36.7%; Pred. No. 6, 5e-21;
 Matches 62; Conservative 24; Mismatches 45; Indels 38; Gaps 3;

QY 25 NKSTQVAVANAPF-----GYTGVIYGVAPLVNDETVKALA 63
 DB 2 NKFTVSLVAVSVAPALAAACSSNNDANGAAGSFGGYS-----VADLQORNT----- 50
 QY 64 SKLPSLVYFDFSDSEIKPQAAALIDEOAFLTTNQTARVLVAGHTDGRSGREYNM 123
 DB 51 -----VYFGPDYKDYDITGEYVQIIDAHAAYINATPAKVLVEGNTDGRGTPPEYN 104
 QY 124 RAVAVRNYLLGKGINQASVETISFGEERPIAGTNEASQRRRAELSY 172
 DB 105 RADAVKGYLAGKGVAGKLGTVSGEERPAVLGHDAAYSKRRRAVLAY 153

RESULT 6
 AAR82947
 ID AAR82947 standard; protein; 153 AA.

AC AAR82947;

PD 10-OCT-1990 (first entry)

DE 16600 dalton outer membrane protein of non-typable H influenzae.

XX Haemophilus influenzae; 16600 dalton outer membrane protein (OMP);

XX pneumonia; meningitis; ss.

OS Haemophilus influenzae.

XX EP281673-A.

PD 14-SEP-1988.

DT 29-JUL-1997 (first entry)
XX
DE H. pylori outer membrane protein 31262.aa.
XX
KW Outer membrane; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bacterium; life cycle; activator;
KW bacterium; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
XX
OS Helicobacter pylori.
XX
PN MO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96MO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR) ASTRA AB.
XX
PI Berglindh OT, Smith D, Møllgaard BL;
XX
DR WPI: 1997-052306/05.
DR N-PSDB; AAT67789.
XX
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 56; Page 577; 1481pp: English.
XX
CC This sequence is a H. pylori outer membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds.
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
XX Sequence 179 AA;
SO
Query Match 25.0%; Score 215.5; DB 18; Length 179;
Best Local Similarity 32.8%; Pred. No. 2.6e-15;
Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;
OY 15 SVLTFM-----TGCANK-----STSOVMVAP--NAPTGYTGVIYGVAPLVND 56
DB 5 SVSFLVAPLLVVGCSHKMDNKTVAADVSTKAVGAPVPTTEAPKEPKQEPAPVVEK 64
OY 57 ETVKALASKLPISLVYFDEPDEIKPQAAAILDEQAQFLTTQTAHVAVAGHTDGRSREY 116
DB 65 PAIE--SGTIIASITFDDPKYEIKSDOETLDEIVQAKENH-MQVLLGNTDEFGSSEY 121
OY 117 NMSLGERRAVAVRNVLGKGINQASVELISFGEEPRPIAFGTNEBAMQONRAEL 170
DB 122 NQALGVKRTLSVKNALVIKGVKDKMKTISFGESKPKCVQKTRCCTYENRRVDV 175

RESULT 12

AAW24651
ID AAW24651 standard; Protein; 179 AA.
XX
AC AAW24651;
XX

DT 11-AUG-1997 (first entry)
XX
DE H. pylori outer membrane protein 31262.aa.
XX
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
XX secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW activator; inhibitor; bacterial life cycle; vaccine; immune;
KW detection; antisense; inhibition.
XX
OS Helicobacter pylori.
XX
PN MO9719098-A1.
XX
PD 29-MAY-1997.
XX
PF 15-NOV-1996; 96MO-US18542.
XX
PR 17-NOV-1995; 95US-0561469.
XX
PA (ASTR) ASTRA AB.
XX
PI Smith DH;
XX
DR WPI: 1997-298052/27.
DR N-PSDB; AAT77469.
XX
XX
PT Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics
XX
PS Claim 18; Page 171; 235pp: English.
XX
CC This sequence represents an H. pylori outer membrane protein.
CC Helicobacter pylori has been strongly linked to chronic gastritis and
CC duodenal ulcer disease. The nucleic acid sequences of the invention
CC are used to evaluate compounds, especially activators or inhibitors of
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC sequence. The nucleic acid sequences, and corresponding proteins, are
CC also useful for generating vaccines for immunising subjects against H.
CC pylori or for use in detecting the presence of Helicobacter species in
CC a sample. Antisense nucleic acid sequences of these sequences are
CC used to inhibit expression of a gene from Helicobacter species. H.
CC pylori whole genomic DNA was isolated and nebulised to a median size of
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
CC complementary to the BstXI-cut pmx vectors, while the overhang is not
CC self-complementary. Therefore the linkers will not concatamerise nor
CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
CC were ligated to each of the 20 pmx vectors to construct a series of
CC shotgun subclone libraries. The purified DNA samples were then
CC sequenced.
CC Note: The ORF/protein reference number for this sequence was obtained
CC from the related specification, WO9640893.
XX
XX Sequence 179 AA;
SO
Query Match 25.0%; Score 215.5; DB 18; Length 179;
Best Local Similarity 32.8%; Pred. No. 2.6e-15;
Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;
OY 15 SVLTFM-----TGCANK-----STSOVMVAP--NAPTGYTGVIYGVAPLVND 56
DB 5 SVSFLVAPLLVVGCSHKMDNKTVAADVSTKAVGAPVPTTEAPKEPKQEPAPVVEK 64
OY 57 ETVKALASKLPISLVYFDEPDEIKPQAAAILDEQAQFLTTQTAHVAVAGHTDGRSREY 116
DB 65 PAIE--SGTIIASITFDDPKYEIKSDOETLDEIVQAKENH-MQVLLGNTDEFGSSEY 121
OY 117 NMSLGERRAVAVRNVLGKGINQASVELISFGEEPRPIAFGTNEBAMQONRAEL 170
DB 122 NQALGVKRTLSVKNALVIKGVKDKMKTISFGESKPKCVQKTRCCTYENRRVDV 175

RESULT 13

ID AAM20795 standard; Protein; 187 AA.

XX AAM20795;

XX 21-JUL-1997 (first entry)

DE H. pylori outer membrane protein 079p31516orf4.

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

OS Helicobacter pylori.

XX MO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96MO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTRA) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI: 1997-052306/05.

XX N-PSDB; AAT68048.

PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
XX infection, and to detect Helicobacter

PS Claim 56; Page 1203; 1481pp; English.

XX This sequence is a H. pylori outer membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. No
CC identity likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
XX production, e.g. in E. coli hosts.

XX Sequence 187 AA;

Query Match 25.0%; Score 215.5; DB 18; Length 187;
Best Local Similarity 32.8%; Pred. No. 2,8e-15;
Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;

QY 15 SVUTM-----TGCANR-----STQVAVP--NAPGTGYITGVAPLVNDND 56

DB 13 SVFSLVAFLLVVGCSHKMNKTVAGDVSTKAVGTAPVTEPAPEKEEPKQEPAPVVEE 72

QY 57 ETVKALASKPLSLVYEDDSDEIKPQAAAILDEQAQFLITQOTARVLYAGHTDERSREY 116

DB 73 PAIE--SGTIIASIFEDDKYEIKESDQETLDEIVQAKENH-MQVLLLEANTDEFGSSSEY 129

QY 117 NMSIGERRAVAVRNVLGKGINQASVETISFGEERPIAFGTNEBAMSONRAEL 170

DB 130 NQALGVKRTLISVKNALVTKGVEKDMITISFGEKPKCAQKTRCYENRNDV 183

RESULT 16
AAM89813
ID AAM89813 standard; Protein; 179 AA.

XX AAM89813;

XX 18-FEB-1999 (first entry)

DE Protein encoded by clone G1a ORF2.

KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

OS Helicobacter pylori.

XX MO9849314-A2.

XX 05-NOV-1998.

XX 27-APR-1998; 98MO-US08487.

XX 14-OCT-1997; 97US-0061958.

XX 25-APR-1997; 97US-0045107.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Chow TP, Fry KE, Lim MY, McAttee CP;

XX WPI: 1999-009433/01.

XX N-PSDB; AAV90545.

PT New Helicobacter pylori antigens and related nucleic acid sequences
PT - useful in serological diagnosis and protective vaccines, providing
XX long-lasting immune response

PS Claim 10; Page 96-97; 402pp; English.

XX The present sequence represents a Helicobacter pylori antigenic protein
CC that is characterised by immunoreactivity with H. pylori positive
CC antisera. The proteins are highly immunogenic and induce a long-lasting
CC immune response that persists even after antimicrobial treatment. In
CC antibody detection assays, on sera, plasma, urine, saliva etc., they are
CC highly sensitive and specific. The specification also describes 69
CC previously unrecognised immunogenic cluster families. H. pylori antigens
CC are used to detect H. pylori-specific antibodies, for diagnosing
CC infection or to confirm eradication of infection, and in vaccines to
CC protect against H. pylori infection and related diseases (gastritis,
XX peptic ulcer, gastric adenocarcinoma/lymphoma).

XX Sequence 179 AA;

Query Match 24.6%; Score 211.5; DB 20; Length 179;
Best Local Similarity 31.2%; Pred. No. 7,2e-15;
Matches 54; Conservative 36; Mismatches 72; Indels 11; Gaps 4;

QY 8 AAAAAALSVLTMTTCANRSTQVAVP--NAPGTGYITGVAPLVNDNDEYKA-LASK 65

DB 4 SASFSLVAFLLVVGCSHKMNKTVAGDVSTKAVGTAPVTEPAPEKEEPKQEPAPVVEE 63

QY 66 LPST-----VYDFPDSDEIKPQAAAILDEQAQFLITQOTARVLYAGHTDERSREY 117

DB 64 KPAISGTTIASIYDFDKYEIKESDQETLDEIVQAKENH-MQVLLLEANTDEFGSSSEY 122

QY 118 MSLGERRAVAVRNVLGKGINQASVETISFGEERPIAFGTNEBAMSONRAEL 170

DB 123 QALGVKRTLISVKNALVTKGVEKDMITISFGEKPKCAQKTRCYENRNDV 175

RESULT 17
AAM23592
ID AAM23592 standard; Protein; 179 AA.
XX AAM23592;
XX 23-JAN-1998 (first entry)

DE H. pylori Omp22 outer membrane protein.
 XX Omp22 gene; outer membrane protein; antigen; immunogen; stomach;
 KW gastritis; peptic ulcers; gastric cancer; diagnostic; vaccine.
 XX
 OS Helicobacter pylori.
 XX
 XX
 FT Key Location/Qualifiers
 FT Peptide 1..35
 FT /label= leader_sequence
 FT Protein 36..179
 FT /label= mature_peptide
 FT /note= "Omp22"
 XX
 XX WO9728264-A1.
 XX
 XX 07-AUG-1997.
 XX
 XX 06-SEP-1996; 96WO-KR00154.
 XX
 XX 30-JAN-1996; 96KR-0002105.
 XX
 XX (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX
 XX Kim J, Moon H, Park Y, Seo W, Yu G, Yum J;
 XX WPI: 1997-402617/37.
 XX N-PSDB: AAT74195.
 DR
 XX Helicobacter pylori outer membrane protein, Omp22 - used in a
 PT vaccine for treatment or prevention of H. pylori infection
 PS
 XX Example 3: Fig 5; 38pp; English.
 CC This sequence represents Omp22, an outer membrane protein which
 CC exhibits antigenicity and immunogenicity against Helicobacter pylori.
 CC This bacterium is associated with inflammation of the stomach and
 CC gastritis related diseases e.g. peptic ulcers and gastric cancer. Omp22
 CC could be used in immunological therapy as a H. pylori-specific antigen
 CC for the treatment and prevention of diseases associated with this
 CC microorganism e.g. as the active ingredient in a diagnostic kit or a
 CC vaccine.
 CC
 XX
 XX Sequence 179 AA:
 SQ
 Query Match 24.2%; Score 208.5; DB 18; Length 179;
 Best Local Similarity 30.8%; Pred. No. 1.5e-14;
 Matches 52; Conservative 33; Mismatches 69; Indels 15; Gaps 4;
 QY 14 LSVLFMTGCMKSTQVMVAF-NAPGTGVIYTG-----APLVNDETVKA 61
 DB 10 LVAFLFIAGCKHNDKREIVAGDVSAKAVQSPVSTELAQKQEPKQEPAPVVEKPAVE- 68
 QY 62 LASKLPISLVFDPDSDEIKPQAAAILDEQAQFLTTNOTARVLYAGHTDEGSEYKMSIG 121
 DB 69 -SGTIIASIFDFDKYEIKESDQETLDEIVQKAKENH-MQVLLGNTDEFGSSSEYQALG 126
 QY 122 ERRANAIVRYLLGKGINASVEIISFGERPARGTNEEAMSONRAEL 170
 DB 127 VRTLSVKNALVIKGVKDMKITISFGETKPKCKQKRECKENRRVDY 175
 RESULT 18
 AAW23591
 ID AAW23591 standard; Protein; 144 AA.
 XX
 XX AAW23591;
 XX
 XX 23-JAN-1998 (first entry)
 XX
 XX H. pylori recombinant Omp22 recombinant outer membrane protein.
 DE
 XX Omp22 gene; outer membrane protein; antigen; immunogen; stomach;
 KW

KW gastritis; peptic ulcers; gastric cancer; diagnostic; vaccine;
 KW recombinant protein.
 XX
 OS Helicobacter pylori.
 XX
 XX
 FT Key Location/Qualifiers
 FT Peptide 1..144
 FT /label= Omp22
 FT /note= "Recombinant outer membrane protein"
 XX
 XX WO9728264-A1.
 XX
 XX 07-AUG-1997.
 XX
 XX 06-SEP-1996; 96WO-KR00154.
 XX
 XX 30-JAN-1996; 96KR-0002105.
 XX
 XX (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 XX
 XX Kim J, Moon H, Park Y, Seo W, Yu G, Yum J;
 XX WPI: 1997-402617/37.
 XX N-PSDB: AAT74194.
 DR
 XX Helicobacter pylori outer membrane protein, Omp22 - used in a
 PT vaccine for treatment or prevention of H. pylori infection
 PS
 XX Claim 2; Pages 25; 38pp; English.
 CC This amino acid sequence of Omp22, is a recombinant outer membrane
 CC protein which is identical to the mature protein of its native form
 CC (see AAW23592). It retains antigenicity and immunogenicity against
 CC Helicobacter pylori. This recombinant Omp22 protein can be mass produced
 CC in via expression in E.coli DH5-alpha cells. H. pylori is associated
 CC with inflammation of the stomach and gastritis related diseases e.g.
 CC peptic ulcers and gastric cancer. The native and the recombinant form of
 CC Omp22 could be used in immunological therapy as a H. pylori-specific
 CC antigen for the treatment and prevention of diseases associated with
 CC this microorganism e.g. as the active ingredient in a diagnostic kit or
 CC a prophylactic/therapeutic vaccine
 CC
 XX
 XX Sequence 144 AA:
 SQ
 Query Match 23.9%; Score 205.5; DB 18; Length 144;
 Best Local Similarity 38.0%; Pred. No. 2.4e-14;
 Matches 46; Conservative 25; Mismatches 47; Indels 3; Gaps 2;
 QY 50 APLVNDDETVKALASKLPISLVFDPDSDEIKPQAAAILDEQAQFLTTNOTARVLYAGHTD 109
 DB 23 APVVEKPAVE--SGTIIASIFDFDKYEIKESDQETLDEIVQKAKENH-MQVLLGNTD 79
 QY 110 ERGSEYKMSIGERRAIVRYLLGKGINASVEIISFGERPARGTNEEAMSONRAEL 169
 DB 80 EFGSEYQALGVRTLSVKNALVIKGVKDMKITISFGETKPKCKQKRECKENRRVDY 139
 QY 170 L 170
 DB 140 V 140
 RESULT 19
 AAY35412
 ID AAY35412 standard; Protein; 192 AA.
 XX
 XX AAY35412;
 XX
 XX 13-SEP-1999 (first entry)
 XX
 XX Chlamydia pneumoniae transmembrane protein sequence.
 DE
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW

KW vaccine; neutralising epitope.
 OS Chlamydia pneumoniae.
 PN WO9227105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 XX 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffiths R;
 XX WPI: 1999-357842/30.
 DR
 XX
 PT
 XX
 PS Genome sequence of Chlamydia pneumoniae
 Page 1196-1197; Disclosure: 1912pp; English.
 XX
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 XX
 SQ Sequence 192 AA;
 Query Match 23.8%; Score 204.5; DB 20; Length 192;
 Best Local Similarity 38.7%; Pred. No. 4.6e-14;
 Matches 46; Conservative 24; Mismatches 46; Indels 3; Gaps 2;
 Oy 54 DNDETVALASKLPSTL--VFDPDSDEIK-POAAATLDEQAQFLTTNQTARVLVAGHTDE 110
 Db 68 DSKEEKQYKSSQVAARNTTFATDSTYIKGEENLALTNLVHYMKKNPKATLYIGGHTDE 127
 Oy 111 RGSREYNMISGERRAVAVRNYLLGKINGASVEIISFGGERPIAGTNEAASQNRRAE 169
 Db 128 RGAASYNLALGARARAAIKERHLEKQGISADRLSTISYGEHPLNSGHNELMAQONRRTE 186
 RESULT 20
 AAE04322
 ID AAE04322 standard; Protein; 192 AA.
 XX
 AC AAE04322;
 XX
 DT 04-SEP-2001 (first entry)
 XX
 DE Chlamydia pneumoniae omp P6 precursor protein.
 XX
 KW Omp P6 precursor; outer membrane protein; Chlamydia infection;
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; atherosclerosis; asthma;
 KW vaccine; antibiotic.
 XX
 OS Chlamydia pneumoniae.
 OS
 XX
 PN WO200146224-A2.
 XX
 PD 28-JUN-2001.
 XX
 PF 20-DEC-2000; 2000WO-CA01534.
 XX

PR 22-DEC-1999; 99US-0171525.
 XX
 PA (AVER) AVENTIS PASTEUR LTD.
 XX
 PI Murdin AD, Oomen RP, Wang J, Dunn P;
 XX
 DR WPI: 2001-408630/43.
 DR N-PSDB; AAD08593.
 XX
 PT Chlamydia polypeptides and polynucleotides for preventing, diagnosing
 XX and treating Chlamydia infection in mammals, especially humans -
 PS Claim 16; Fig 1; 74pp; English.
 XX
 CC The present sequence is Chlamydia pneumoniae outer membrane
 CC protein (omp) P6 precursor. The omp P6 precursor polynucleotide,
 CC polypeptide and its antibody are useful for detecting, preventing and
 CC treating Chlamydia infections e.g. pneumonia, upper respiratory tract
 CC diseases (e.g. bronchitis, sinusitis) and acute respiratory diseases
 CC (e.g. cough, sore throat). C. pneumoniae infection is also observed to
 CC be associated with atherosclerosis and asthma. Immunogenic fragments
 CC of omp P6 precursor protein and their corresponding DNAs are useful in
 CC vaccine composition. The polynucleotide is useful in the construction
 CC of attenuated Chlamydia strains that can over express the polynucleotide
 CC or express it in a non-toxic, mutated form. The omp P6 probes are useful
 CC in diagnostic tests as capture or detection probes and the primers are
 CC useful in diagnostic methods involving PCR (polymerase chain reaction).
 CC The antibody is useful for purifying polypeptide.
 CC
 XX
 SQ Sequence 192 AA;
 Query Match 23.8%; Score 204.5; DB 22; Length 192;
 Best Local Similarity 38.7%; Pred. No. 4.6e-14;
 Matches 46; Conservative 24; Mismatches 46; Indels 3; Gaps 2;
 Oy 54 DNDETVALASKLPSTL--VFDPDSDEIK-POAAATLDEQAQFLTTNQTARVLVAGHTDE 110
 Db 68 DSKEEKQYKSSQVAARNTTFATDSTYIKGEENLALTNLVHYMKKNPKATLYIGGHTDE 127
 Oy 111 RGSREYNMISGERRAVAVRNYLLGKINGASVEIISFGGERPIAGTNEAASQNRRAE 169
 Db 128 RGAASYNLALGARARAAIKERHLEKQGISADRLSTISYGEHPLNSGHNELMAQONRRTE 186
 RESULT 21
 ABB90544
 ID ABB90544 standard; Protein; 192 AA.
 XX
 AC ABB90544;
 XX
 DT 29-JUL-2002 (first entry)
 XX
 DE Chlamydia pneumoniae cp7090 protein, SEQ ID NO:37.
 XX
 KW Chlamydia infection; antigen; immunogen; vaccine; diagnosis;
 KW human respiratory disease; cardiovascular disease; atherosclerosis;
 KW coronary artery disease; carotid artery stenosis; myocardial infarction;
 KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
 KW strain CWL029.
 XX
 OS Chlamydia pneumoniae.
 OS
 XX
 FH key
 FT Peptide 1..20
 FT /label= signal_peptide
 FT 21..192
 FT /note= "Mature protein"
 XX
 PN WO200202606-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 03-JUL-2001; 2001WO-IB01445.
 XX

XX 03-JUL-2000; 2000GB-0016363.
 PR 11-JUL-2000; 2000GB-0017047.
 PR 21-JUL-2000; 2000GB-0017983.
 PR 07-AUG-2000; 2000GB-0019368.
 PR 18-AUG-2000; 2000GB-0020440.
 PR 14-SEP-2000; 2000GB-0022583.
 PR 10-NOV-2000; 2000GB-0027549.
 PR 22-DEC-2000; 2000GB-0031706.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Ratti G, Grandi G;
 XX
 DR WPI: 2002-154726/20.
 DR N-PSDB; ABL91202.
 XX
 PT Novel Chlamydia pneumoniae protein useful in the manufacture of a
 PT medicament for treatment or prevention of infection due to Chlamydia,
 PT preferably Chlamydia pneumoniae, and for diagnostic purposes
 PS Claim 1; Page 60; 364pp; English.
 XX
 CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
 CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
 CC them. The proteins are predicted to be immunogenic and may therefore be
 CC useful in vaccine production and for diagnostic purposes. Chlamydia
 CC pneumoniae is a common cause of respiratory disease in humans, and is
 CC also involved in the development of cardiovascular diseases such as
 CC atherosclerosis, coronary artery disease, carotid artery stenosis,
 CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
 CC claudication and stroke. The proteins and nucleic acids of the invention
 CC may be used in vaccines and pharmaceutical compositions for the
 CC prevention or treatment of chlamydial infections, particularly Chlamydia
 CC pneumoniae infections. The proteins may also be used in the detection of
 CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
 CC DNA probe assay or blotting techniques for determining Chlamydia
 CC pneumoniae gene expression. The present sequence represents a
 CC specifically claimed Chlamydia pneumoniae protein of the invention.
 CC
 XX
 SQ Sequence 192 AA;
 Query Match 23.8%; Score 204.5; DB 23; Length 192;
 Best Local Similarity 38.7%; Pred. No. 4.6e-14;
 Matches 46; Conservative 24; Mismatches 46; Indels 3; Gaps 2;
 QY 54 DNDFTVKALASKLPSEL--VFEDPDSDEIK-PQAAIIDEQAQFLTTQOTARVLVAGHTDE 110
 DB 68 DSKEKQYKSSQVAAFNNITFATDSTYTKGENTLILNLVHYKKPKATLYIEGHTDE 127
 QY 111 RGSREYNMSLGERRAVAVANNYLKGINQASVEIISFGEERPIAFGTNEAMSQRRAE 169
 DB 128 RGAASYNLALGARRRNAIKHELRKQGISADRLSTSYGKEHPHLSNGHNEALMOQNRTE 186

RESULT 22
 AAM89983
 ID AAM89983 standard; Protein; 104 AA.
 AC AAM89983;
 XX
 DT 18-FEB-1999 (first entry)
 XX
 DE Expressed antigen for clone Y89A.
 XX
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
 OS Helicobacter pylori.
 XX
 PN WC9849314-A2.
 XX
 PD 05-NOV-1998.

XX 27-APR-1998; 98WO-US08487.
 PF
 XX 14-OCT-1997; 97US-0061958.
 XX 25-APR-1997; 97US-0045107.
 PR
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 XX
 PI Chow TP, Fry KE, Lim MY, McAtee CP;
 XX
 DR WPI: 1999-009433/01.
 XX
 PT New Helicobacter pylori antigens and related nucleic acid sequences
 PT - useful in serological diagnosis and protective vaccines, providing
 PT long-lasting immune response
 PS Claim 16; Page 322; 402pp; English.
 XX
 CC The present sequence represents a Helicobacter pylori antigenic protein
 CC that is characterised by immunoreactivity with H. pylori-positive
 CC antisera. The proteins are highly immunogenic and induce a long-lasting
 CC immune response that persists even after antimicrobial treatment. In
 CC antibody-detection assays, on sera, plasma, urine, saliva etc., they are
 CC highly sensitive and specific. The specification also describes 69
 CC previously unrecognised immunogenic cluster families. H. pylori antigens
 CC are used to detect H. pylori-specific antibodies, for diagnosing
 CC infection or to confirm eradication of infection, and in vaccines to
 CC protect against H. pylori infection and related diseases (gastritis,
 CC peptic ulcer, gastric adenocarcinoma/lymphoma).
 CC
 XX
 SQ Sequence 104 AA;
 Query Match 22.9%; Score 197.5; DB 20; Length 104;
 Best Local Similarity 41.6%; Pred. No. 1.1e-13;
 Matches 42; Conservative 19; Mismatches 39; Indels 1; Gaps 1;
 QY 70 VFPEFSDDEIKPQAAIIDEQAQFLTTQOTARVLVAGHTDEGSRREYNMSLGERRAVAVR 129
 DB 1 IYFPDKEIKESQETLDELIVQAKENH-MQVLLGNGDERSSEYNOALGVKRTLSVK 59
 QY 130 NYLLGKGINQASVEIISFGEERPIAFGTNEAMSQRRAE 170
 DB 60 NALVIKGVKMDIKTISFGETKPKCAQKRECYKRNRRDV 100

RESULT 23
 AAY37603
 ID AAY37603 standard; Protein; 214 AA.
 AC AAY37603;
 XX
 DT 07-OCT-1999 (first entry)
 XX
 DE Chlamydia trachomatis surface exposed protein.
 XX
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; peritropatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW bartolinitis; pneumopathy; venereal lymphogranulomatosis.
 XX
 OS Chlamydia trachomatis.
 XX
 PN WC9928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 DT 27-NOV-1998; 98WO-IB01939.
 PF
 XX 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GEST) GENSET.

XX Griffiths R;
XX WPI; 1999-371125/31.
XX
XX Genome sequence of Chlamydia trachomatis
XX
XX Disclosure; Page 1248-1249; 1755pp; English.
XX
XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
XX of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
XX vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
XX can also be used to control growth of the microorganism. Chlamydia
XX trachomatis is responsible for a large number of diseases, e.g. eye
XX diseases such as conventional trachoma, nongonococcal urethritis,
XX paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX perleptitis, bartonellosis, pneumopathy in breast feeding infants;
XX and venereal lymphogranulomatosis. The polypeptides of the invention
XX may be of use in treating these diseases.

SO Sequence 214 AA;

Query Match 22.8%; Score 196.5; DB 20; Length 214;
Best Local Similarity 35.3%; Pred. No. 4e-13;
Matches 49; Conservative 19; Mismatches 54; Indels 17; Gaps 2;

QY 48 GVAFLVNDERTVAKASKLPSP-----VFEDPSDEIKRQ-AAATLDEQ 90
DB 70 GFVFPYSDEEIQAFVDFDSKEBQLYKTSAGQSPFNITFADTSIKGEDNLTTLASL 129
QY 91 AQTFTNOTAVLVAAGHTDERSREYMSLGERRAVAARNYLLGKGINQASVEIISFGE 150
DB 130 VRHLSKPKATLYIEGHTDERGAAYNALGARRAVAAYVQYLKQGIADRLFTISYCKE 189
QY 151 RPIAFGTNEAWSONRRAE 169
DB 190 HPVHPGHNELMAOONRRTE 208

RESULT 24

AAB20105
ID AAB20105 standard; Protein; 224 AA.

XX AAB20105;

XX 23-APR-2001 (first entry)

XX Moraxella catarrhalis BASB113 protein.

XX BASB113; infection; otitis media; pneumonia; therapy; diagnosis;

XX antibacterial; antimicrobial.

XX Moraxella catarrhalis.

XX WO200100836-A1.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-EP05851.

XX 25-JUN-1999; 99GB-0015044.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonard J;

XX WPI; 2001-112458/12.

XX N-PSDB: AAF30043.

XX New BASB113 polypeptide isolated from Moraxella catarrhalis bacterium,
XX useful for diagnosing and producing vaccines against bacterial
XX infections such as otitis media and pneumonia

XX Claim 1; Page 67; 86pp; English.
XX
XX

XX The present sequence is that of BASB113 protein from Moraxella
XX catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
XX media in children and pneumonia in adults. The invention provides
XX BASB113 polypeptides, and polynucleotides encoding them, as well as
XX expression vectors, host cells and methods for producing BASB113
XX polypeptides using recombinant methods. Also claimed is a vaccine
XX composition comprising a BASB113 polypeptide, an immunogenic
XX fragment of a BASB113 polypeptide, or a polypeptide having at least
XX 85% amino acid sequence identity to BASB113, or comprising a
XX polynucleotide encoding such a polypeptide. A claimed method of
XX diagnosing a Moraxella infection involves identifying a BASB113
XX polypeptide or antibody. A claimed therapeutic composition useful
XX in treating humans with M. catarrhalis infection comprises at least
XX 1 antibody directed against a BASB113 polypeptide. BASB113
XX polypeptides also have utility in raising specific antibodies,
XX and in screening for antibacterial drugs.

SO Sequence 224 AA;

Query Match 21.0%; Score 180.5; DB 22; Length 224;
Best Local Similarity 27.0%; Pred. No. 2.4e-11;
Matches 48; Conservative 25; Mismatches 88; Indels 17; Gaps 3;

QY 7 IAAAAALSVLTFTMGCAKSTSOVWVAPNAPGYGVY-----TGAPL 52
DB 42 IGAVAGALGGTALSKATGKEKTRGAIL-GAAYGAAGAAMERQAKIQDMQGGYGT-- 98
QY 53 VDNDEYKALASKIPSLVYDFEDPSDEIKPQAAALDEQAQFTTNOTAVLVAAGHTDERG 112
DB 99 VTHDTGNTLMPGNTFPAHDDTLNSAFLGRNLQNLANTMNQHEHTIYVAGHTDSTG 158
QY 113 SREYMSLGERRAVAARNYLLGKGINQASVEIISFGEERPIAFGTNEAWSONRRAE 170
DB 159 QAAYNQELSRRAVRYIYLINGQVDPYRIQTVGGMKPIASNTEAGRQNRREVEL 216

RESULT 25

AAB47447
ID AAB47447 standard; Protein; 353 AA.

XX AAB47447;

XX 31-OCT-2001 (first entry)

XX MOMP P5.

XX surface exposed loop; major outer membrane protein P5; MOMP P5;

XX non-typable H. influenzae; nH; LBI(f) peptide; B cell epitope;

XX lower respiratory tract infection.

XX Haemophilus influenzae.

XX Key Location/Qualifiers

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX Domain

XX WO200161013-A1.

XX 23-AUG-2001.

XX 13-FEB-2001; 2001MO-EP01556.
 PF 15-FEB-2000; 2000GB-0003502.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA Berthel FJ, Denoel P, Poolman J, Thonnard J;
 PI WPI; 2001-522599/57.
 XX
 DR Recombinant bacterial outer membrane protein where one or more
 PT surface-exposed loops are modified is useful as a vaccine to prevent or
 PT treat Haemophilus Influenzae infection or associated disease, e.g.,
 PT otitis media and conjunctivitis -
 PS
 XX Disclosure; Fig 1; 29pp; English.
 PS This sequence represents the major outer membrane protein P5 of
 CC non-typeable H. Influenzae. One or more surface exposed loops of this
 CC protein may be replaced with a modified peptide of the invention. Each
 CC of these peptides contain an Lbl(f) peptide which is a 19 amino acid
 CC peptide derived from the sequence of MOMP P5 from strain nH1128,
 CC representing amino acids Arg117 to Gly135. This peptide represents the
 CC third exposed loop of P5 and is a potential B cell epitope. The loops
 CC of the invention are modified in terms of being in a non-native
 CC environment in the recombinant outer membrane protein. The modified
 CC MOMP P5 may be used to induce an immune response in a mammal to
 CC prevent or treat Haemophilus influenzae infection or associated
 CC disease, e.g., otitis media, sinusitis, conjunctivitis, or lower
 CC respiratory tract infection.
 CC
 XX
 SQ Sequence 353 AA;
 Query Match 20.5%; Score 176.5; DB 22; Length 353;
 Best Local Similarity 30.6%; Pred. No. 1.2e-10;
 Matches 48; Conservative 27; Mismatches 61; Indels 21; Gaps 4;
 QY 35 PNAPTGY-----TGVIY--TGAPLVNDDETAKALSKLPFLVYFDDSDKRPQA 83
 DB 193 PNTALNTNWMISINAGISYRGGCAAPVAAPVAVKTSFS-LNSDYFAFGKANKLPQA 251
 QY 84 AAILDEQAFLTTNOTARYLVAGHTDEGSRREYNMSLGERRAVAVRYLLGKGINQASVE 143
 DB 252 QATLDSIYGEQMSQVAKVAVAGYDRIGSDAFENKLSQERADSVANFEVAKGVADATIS 311
 QY 144 IISFGERPIAFGTNE-----AMSONRAELS 171
 DB 312 ATGYGRANFVTGATCDQVNGKRALIACRPDRRVEIA 348
 RESULT 26
 AAR99626
 ID AAR99626 standard; Protein; 214 AA.
 XX
 AC AAR99626;
 XX
 DT 01-OCT-1996 (first entry)
 XX
 DE P. aeruginosa orpF-oprI fusion protein.
 XX
 OS Vaccine; fusion protein; OprF; OprI; outer membrane protein;
 KW antibody; glutathione-S-transferase; GST.
 XX
 OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).
 XX
 FH Key Location/Qualifiers
 FT 1..151
 FT Protein /label= oprf
 FT /note= "oprF amino acids 192-342"
 FT 152..214
 FT Protein /label= oprI
 FT /note= "oprI amino acids 21-83"

XX
 PN EF17106-A1.
 XX
 PD 19-JUN-1996.
 XX
 XX 17-NOV-1995; 95BP-0118098.
 PF 16-DEC-1994; 94BP-0120023.
 XX
 XX (BEHM) BEHRINGERWERKE AG.
 PA Broeker M, Dondely H, Hungerer K, Knapp B, Ulrich B;
 PI Von Specht BU;
 PI WPI; 1996-279559/29.
 DR N-PSDB; AAT32600.
 XX
 XX Fusion proteins for Pseudomonas aeruginosa vaccines - contg.
 PT fragments of outer membrane proteins I and F.
 PS
 XX
 PS Disclosure; Page 13-14; 23pp; English.
 PS A cDNA sequence (AAT32600) codes for a fusion (AAR99626) between amino
 CC acids 192-342 of the outer membrane protein F (oprF) and amino acids
 CC 21-83 of outer membrane protein I (oprI) of Pseudomonas aeruginosa
 CC ATCC 33354. A glutathione-S-transferase fusion with the hybrid
 CC protein was expressed in Escherichia coli and Saccharomyces cerevisiae
 CC transformants. The oprF-oprI hybrid protein, and antibodies raised
 CC against it, conferred protection against P. aeruginosa infection in
 CC laboratory animals. The hybrid protein was significantly more
 CC immunogenic than an oprI-oprF fusion (AAR99627).
 CC
 XX
 SQ Sequence 214 AA;
 Query Match 20.4%; Score 175.5; DB 17; Length 214;
 Best Local Similarity 37.5%; Pred. No. 7.8e-11;
 Matches 39; Conservative 16; Mismatches 48; Indels 1; Gaps 1;
 QY 70 VFDDSDSDKRPQAAILDEQAFLTTNOTARYLVAGHTDEGSRREYNMSLGERRAVAVR 129
 DB 51 VFDDSDKRVKSNSTADIKNLADPMKQYPSISTVBEHTSDVGTADVANKLSRRANAVR 110
 QY 130 NYLLGK-GINQASVELISFGERPIAFGTNEAMSONRAELST 172
 DB 111 DVLVNEYGVEGGRVNAVGYGSRPVADNATAGRAINRRVSSH 154
 RESULT 27
 AAR85450
 ID AAR85450 standard; Protein; 338 AA.
 XX
 AC AAR85450;
 XX
 DT 15-FEB-1996 (first entry)
 XX
 DE Nontypable H. Influenzae P5 protein.
 XX
 OS P5 outer membrane protein; vaccine; otitis media; sinusitis;
 KW chronic pulmonary obstructive disease.
 XX
 OS Haemophilus Influenzae.
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 195
 FT /note= "amino acid at position 195 is not
 FT identified in the specification"
 FT MISC-difference 311
 FT /note= "amino acid at position 311 is not
 FT identified in the specification"
 FT
 FT EP680765-A1.
 XX
 PD 08-NOV-1995.

XX 02-MAY-1995; 95EP-0302996.
 XX 05-MAY-1994; 94US-0210394.
 XX (AMCY) AMERICAN CYANAMID CO.
 XX Zlotnick GW;
 XX WPI; 1995-375029/49.
 XX Purified H. influenzae P5 outer membrane protein - used for
 XX preventing reducing susceptibility to or treating H. influenzae
 XX infections
 XX Disclosure; Page 7-8; 16pp; English.
 XX CC Montypable H. influenzae HI outer membrane protein P5 was isolated
 XX CC by extraction of the outer membrane with detergents and cation-exchange
 XX CC chromatography. P5 (or its peptide fragments) are used in vaccines
 XX CC for prevention of H. influenzae infections implicated in otitis media,
 XX CC sinusitis and chronic pulmonary obstructive disease.
 XX SQ Sequence 338 AA;
 XX
 XX Query Match 20.3%; Score 174.5; DB 16; Length 338;
 XX Best Local Similarity 28.9%; Pred. No. 1.9e-10;
 XX Matches 46; Conservative 27; Mismatches 57; Indels 29; Gaps 5;
 XX
 XX QY 33 VAPNAPTGY-----TGVIY---TGVAPLVNDDEYKALASLPISLYVFDPSDEIKP 81
 XX DB 184 IMPNTAIHNPXIGSINNGISYRFGGAP-----VKTFSLND--VFPAFGKANKLP 234
 XX QY 82 QAAATLIDBQAQFLTNTQARVAVAGHTDGRSGREYMSLGERRAVAVRNLGKGINQAS 141
 XX DB 235 QAAATLIDBQAGMSQVSKAKVAVAGYTDRTGSDAFNFKLSQERADSVANFVAKGVAAADA 234
 XX QY 142 VEISFGEERPIAFGTNEAMSQ-----NNRAELSL 171
 XX DB 295 ISATGYGRKANPVTGATNDQVGRMALIATLAPDRREIA 333
 XX
 XX RESULT 28
 XX AAB59180
 XX ID AAB59180 standard; protein; 326 AA.
 XX AC AAB59180;
 XX DT 22-MAR-2001 (first entry)
 XX DE Protein associated with C.coli and C.jejuni CadF.
 XX KW Fibronectin binding protein; CadF; vaccine; diagnostic assay.
 XX OS Pseudomonas fluorescens.
 XX PN US6156546-A.
 XX PD 05-DEC-2000.
 XX PF 15-MAY-1998; 98US-0080025.
 XX PR 16-MAY-1997; 97US-0046763.
 XX PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
 XX PI Garvis SG, Konkel ME;
 XX DR WPI; 2001-079546/09.
 XX PT Novel isolated polynucleotide useful for producing fibronectin binding
 XX PT proteins which are useful in production of vaccine, in diagnostic
 XX PT assays and for prophylactic and therapeutic purposes -

XX Disclosure; Column 39-42; 29pp; English.
 XX PS The present invention relates to a Campylobacter jejuni or
 XX CC Campylobacter coli fibronectin binding protein (CadF). A recombinant
 XX CC expression vector with cadF is useful in an assay for determining the
 XX CC presence of C.jejuni or C.coli in a test sample or for
 XX CC determining whether a test isolate of Campylobacter is a strain
 XX CC of C.coli. cadF is useful in the construction of DNA probes for
 XX CC identifying and quantifying the level of expression of CadF in a
 XX CC cell. The gene can also be used in a vaccine.
 XX SQ Sequence 326 AA;
 XX
 XX Query Match 20.2%; Score 174; DB 22; Length 326;
 XX Best Local Similarity 36.3%; Pred. No. 2e-10;
 XX Matches 37; Conservative 14; Mismatches 51; Indels 0; Gaps 0;
 XX
 XX QY 70 VFDFPSDEIKPQAAIILDEQAQFLTNTQARVAVAGHTDGRSGREYMSLGERRAVAVR 129
 XX DB 219 VKFDFKSVKKPNSYGDVKNLADFMAYPATRNEVAGHTDSTGPDAYNQLSRADRVK 278
 XX QY 130 NYILGKGINQASVEIISGCEERPIAFGTNEAMSQNNRAELSL 171
 XX DB 279 QVLVKDGVAPSRITAVGGERPVDNATEAGRAVNRREVAS 320
 XX
 XX RESULT 29
 XX AAR99625
 XX ID AAR99625 standard; protein; 161 AA.
 XX AC AAR99625;
 XX DT 01-OCT-1996 (first entry)
 XX DE P. aeruginosa oprF C-terminal region (aa190-350).
 XX KW Vaccine: fusion protein; OprF; OprI; outer membrane protein;
 XX KW antibody; glutathione-S-transferase; GST.
 XX OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).
 XX PN EP717106-A1.
 XX PD 19-JUN-1996.
 XX PF 17-NOV-1995; 95EP-0118098.
 XX PR 16-DEC-1994; 94EP-0120023.
 XX PA (BEHM) BEHRINGWERKE AG.
 XX PI Broecker M, Domdey H, Hungerer K, Knapp B, Ulrich B;
 XX PI Von Specht BU;
 XX DR WPI; 1996-279559/29.
 XX DR N-PSDB; AAT132599.
 XX PT Fusion proteins for Pseudomonas aeruginosa vaccines - contg.
 XX PT fragments of outer membrane proteins I and F.
 XX PS Disclosure; Page 11-12; 23pp; English.
 XX CC A cDNA sequence (AAT132599) codes for the C-terminal region, amino
 XX CC acids 190-350 (AAR99625), of the Pseudomonas aeruginosa ATCC 33354
 XX CC outer membrane protein F (oprF). A hybrid protein between
 XX CC glutathione-S-transferase and the oprF moiety was expressed in
 XX CC Escherichia coli. OprF-OprI (AAR99626) and OprI-OprF (AAR99627) fusion
 XX CC proteins were also prep. and tested for their efficacy as vaccines
 XX CC against P. aeruginosa infection.
 XX SQ Sequence 161 AA;

```

Oy      70  VYDFPDSDEIKPQAAALIDEOAOFLLTNTOTARVLYAGHTDEGSEYEMMSLGERRAVVR 129
        | 1111 : : | : : | : : | 1111 : : | 1111 111
Db      118 VKDFEDFSKYEKENSYADIKRLADFMKQYPTSTVEGHNDVSQTDAYNOKLSERRANAVR 177
        | : : : : | : : | : : | : : | : : | : : |
Oy      130 NYLLGK-GINQASVEIISFGEERPIAFGTIEEAMSONRRAE 169
        : : : : | : : | : : | : : | : : | : : |
Db      302 DVLVNEYGVGEGRVANAVGYGESRPVADNMTAEGRAINRRVE 342
        : : : : | : : | : : | : : | : : | : : |

RESULT 31
ID      AAP82053
        AAP82053 standard; protein: 350 AA.
XX      AAP82053;
XX      22-OCT-1990 (first entry)
DE      Outer membrane protein F of Pseudomonas aeruginosa.
XX      Outer membrane protein F; OMPF; vaccination; antibodies.
XX      Pseudomonas aeruginosa.
XX      DE3718591-A.
XX      15-DEC-1987.
PD      03-JUN-1987; 87DE-3718591.
XX      03-JUN-1987; 87DE-3718591.
XX      03-JUN-1987; 87DE-3718591.
XX      (BEHM ) BEHRINGWERKE AG.
PA      Domdey H, Lottspeich F, von Specht B-U, Duchene M;
PI      WPI; 1988-361619/51.
DR      N-PSDB; AAN82023.
XX      New outer membrane protein F of Pseudomonas aeruginosa -
PT      DNA sequences encoding it and derived antibodies, useful for
PT      vaccination and diagnosis.
PS      Disclosure; ; P; German.
XX      The protein is isolated from the OMP of P. aeruginosa serotype 6
CC      ATTC 33354) and purified by HPLC. The amino-terminal and trypsin
CC      fragments are sequenced and a series of oligonucleotide probes
CC      constructed corresponding to the established sequences. These probes are
CC      used to screen a gene bank of 15-20 kb fragments of genomic DNA in
CC      lambda EMBL 3. One positive clone includes a 2.5 kb PstI fragment. This
CC      protein gene, which can be isolated as a high copy no. vector because of the
CC      fragment cannot be cloned int a high copy no. vector because of the
CC      toxicity of the gene prod., so is subcloned as two fragments with an
CC      overlapping region of about 500bp. Ab's are raised by usual immunisation
CC      or cell-fusion procedures. The Ab's are useful in diagnosis.
XX      Sequence 350 AA:
SQ
Query Match 19.9%; Score 171.5; DB 9; Length 350;
Best Local Similarity 37.6%; Pred. No. 4,2e-10;
Matches 38; Conservative 15; Mismatches 47; Indels 1; Gaps 1;

Oy      70  VYDFPDSDEIKPQAAALIDEOAOFLLTNTOTARVLYAGHTDEGSEYEMMSLGERRAVVR 129
        | 1111 : : | : : | : : | 1111 : : | 1111 111
Db      242 VKDFEDFSKYEKENSYADIKRLADFMKQYPTSTVEGHNDVSQTDAYNOKLSERRANAVR 301
        | : : : : | : : | : : | : : | : : | : : |
Oy      130 NYLLGK-GINQASVEIISFGEERPIAFGTIEEAMSONRRAE 169
        : : : : | : : | : : | : : | : : | : : |
Db      302 DVLVNEYGVGEGRVANAVGYGESRPVADNMTAEGRAINRRVE 342
        : : : : | : : | : : | : : | : : | : : |

RESULT 32

```

AA859179
ID AA859179 standard; protein; 350 AA.
XX
AC AA859179;
XX
DT 22-MAR-2001 (first entry)
XX
DE Protein associated with C.coli and C.jejuni Cadf.
XX
KW Fibronectin binding protein; Cadf; vaccine; diagnostic assay.
XX
OS Pseudomonas aeruginosa.
XX
PN US6156546-A.
XX
PD 05-DEC-2000.
XX
PF 15-MAY-1998; 98US-0080025.
XX
PR 16-MAY-1997; 97US-0046763.
XX
PA (UNIV) UNIV WASHINGTON STATE RES FOUND.
XX
PI Garvis SG, Konkel ME;
XX
DR WPI; 2001-079546/09.
XX
PT Novel isolated polynucleotide useful for producing fibronectin binding
PT proteins which are useful in production of vaccine, in diagnostic
XX assays and for prophylactic and therapeutic purposes
XX
PS Disclosure; Column 37-40; 29pp; English.
XX
CC The present invention relates to a Campylobacter jejuni or
CC Campylobacter coli fibronectin binding protein (Cadf). A recombinant
CC expression vector with cadf is useful in an assay for determining the
CC presence of C.jejuni or C.coli in a test sample or for
CC determining whether a test isolate of Campylobacter is a strain
CC of C.coli. cadf is useful in the construction of DNA probes for
CC identifying and quantifying the level of expression of Cadf in a
CC cell. The gene can also be used in a vaccine.
XX
SQ Sequence 350 AA;
XX
Query Match 19.9%; Score 171.5; DB 22; Length 350;
Best Local Similarity 37.6%; Pred. No. 4.2e-10;
Matches 38; Conservative 15; Mismatches 47; Indels 1; Gaps 1;
XX
QY 70 YFFDDSEIKPQAAAIIDQAOFLTTNQTARVLVAGHTDGRGSRREYNMISLGRRAVAYR 129
DB 242 VKFEDKSKVKENSVADIKNLADPMKQYPTSTVEGHTDSVGTDAYNOKLSERRANAVR 301
QY 130 NYLTKG-GINQASVEIISFGERPPIAFCTNEAMSONRAE 169
DB 302 DVLVNEKVEGGRYNAVGVGSRFVADNATAGEGRAINRVE 342
XX
RESULT 33
AA96098
ID AA96098 standard; Protein; 369 AA.
XX
AC AA96098;
XX
DT 19-DEC-2000 (first entry)
XX
DE Actinobacillus pleuropneumoniae OmpA2.
XX
KW Outer membrane protein; OmpA; vaccine; pig; immunogen;
KW dual immune response; immunogen; pleuropneumonia.
OS Actinobacillus pleuropneumoniae.
XX
PN EPI035133-A2.

XX
XX 13-SEP-2000.
XX
PF 14-FEB-2000; 2000EP-0301103.
XX
PR 17-FEB-1999; 99US-0120454.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Campos M, Martind SR, Durtschl BA, Yule TD;
XX
DR WPI; 2000-566924/53.
XX
PT Novel fusion protein for producing a dual immune response comprises a
PT peptide analogous to an endogenous peptide which is to be inhibited,
PT connected to a peptide analogous to an immunogen from a pathogen which
XX infects a vertebrate
XX
PS Disclosure; Page 58-60; 93pp; English.
XX
CC The present sequence is that of outer membrane protein OmpA2 of
CC Actinobacillus pleuropneumoniae. OmpA2 can be utilised in fusion
CC proteins of the invention that are used as vaccines eliciting a
CC dual immune response. Such fusion proteins comprise: a first
CC protein that is endogenous to a vertebrate, the activity of which
CC is to be inhibited within the vertebrate, and which is incapable of
CC itself of eliciting an effective immunoinhibitory response in the
CC vertebrate; and a second protein, which is an immunogen from a
CC pathogen capable of infecting the vertebrate, and which causes the
CC vertebrate's immune system to recognise the first protein,
CC producing a response that inhibits the activity of the first
CC protein, and also protecting the vertebrate from infection by the
CC pathogen when the vertebrate is vaccinated with the fusion protein.
CC In the present case, a fusion protein of cholestyrolin and OmpA2
CC is useful for encouraging appetite in swine while simultaneously
CC providing a protective immune response against porcine
CC pleuropneumonia.
XX
SQ Sequence 369 AA;
XX
Query Match 19.7%; Score 170; DB 21; Length 369;
Best Local Similarity 34.5%; Pred. No. 6.6e-10;
Matches 49; Conservative 20; Mismatches 61; Indels 12; Gaps 5;
XX
QY 34 APNAPTGTGYIYI---GVALVNDDEYKALASKLPSLVYDFDSEIKPQAAAIIDE 89
DB 210 APDIHSYVYAGLSYRFGGAVAPVPEPVTKNFA--FSSDYLFDFKSSIKPAAATALD- 266
QY 90 QAOELT---QTARVLVAGHTDGRGSRREYNMISLGRRAVAVANNYLIGGINQASVEIIS 146
DB 267 AANTEIANLGIATPAIDVYNGYITDRIGKASNLKSORRAETVANNYLVSQGNPANNYAVG 326
QY 147 FGERPIAFCTNEAMSONRAE 168
DB 327 YGEANPVYGTATCDKV--KGRKA 346
XX
RESULT 34
AA97900
ID AA97900 standard; Protein; 369 AA.
XX
AC AA97900;
XX
DT 11-SEP-2000 (first entry)
XX
DE Actinobacillus pleuropneumoniae outer membrane protein, OmpA2.
XX
KW OmpA2; outer membrane protein; APP; pneumonic pathogen; swine;
KW Sus scrofa; serotype; antigen; vaccine; cross-reactive immunity;
KW cross-protection; diagnosis.
XX
OS Actinobacillus pleuropneumoniae strain Pz420.
XX

FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /note="Signal peptide"
 FT Protein 20..369
 FT /note="Mature OmpA2"
 XX
 FN EPI001025-A2.
 XX
 PD 17-MAY-2000.
 XX
 PF 20-OCT-1999; 99EP-0308262.
 XX
 PR 22-OCT-1998; 98US-0105285.
 XX
 PA (PE12) PEIZER PROD INC.
 XX
 PI Ankenbauer RG, Baarsch MJ, Campos M, Keich RL, Rosey EL;
 PI Warren-Stewart LM, Sulter BT;
 XX
 DR WPI: 2000-320438/28.
 DR N-PSDB: AAA38558.
 XX
 PT Low molecular weight Actinobacillus pleuropneumoniae proteins and DNA
 PT encoding them, for use as vaccines against the bacteria in swine
 XX
 PS Claim 2; Fig 5; 81pp; English.
 XX
 CC This sequence represents a low molecular weight outer membrane
 CC protein, OmpA2, from Actinobacillus pleuropneumoniae (AP) strain
 CC P2420 (ATCC 98930). The invention relates to the novel AP outer
 CC membrane proteins Omp20, Omp21, Omp22, Omp23, Omp24, Omp25, Omp26
 CC and to nucleic acids encoding them (AAA38554-38558). AP is a Gram
 CC negative coccobacillus which is one of the most important swine
 CC pneumonia pathogens. 12 different serotypes of AP have been recognised
 CC which vary in geographic distribution. Prior art attempts at vaccinating
 CC against AP have produced mainly serotype-specific immune responses. In
 CC contrast, natural immunity to any one serotype seems to confer
 CC significant protection from disease caused by other serotypes, suggesting
 CC that natural outer membrane proteins of the invention are present in all 12
 CC serotypes, and may provide a target for cross-protective immunisation.
 CC The novel outer membrane proteins and nucleic acids encoding them can be
 CC used as a vaccine against AP in swine. They can also be used as
 CC reagents for the diagnosis of AP infections.
 CC
 SQ Sequence 369 AA:
 Query Match 19.7%; Score 170; DB 21; Length 369;
 Best Local Similarity 34.5%; Pred. No. 6, 6e-10;
 Matches 49; Conservative 20; Mismatches 61; Indels 12; Gaps 5;
 OY 34 APNAPGTYGVYIT---GVAPLVNDDETVKALSKLPISLYVDFDSDEIKPOAAALIDE 89
 DB 210 APDIHSTVAGLSYRFGQAVAPVEPEVYTKNFA--FSSDYLFDFGKSSLPAAATALD- 266
 OY 90 QAOFLITN--QTARVLYAGHTDERSREYMSIGERRAAVAVRYTLGKGINQASVEIIS 146
 DB 267 AANTEIANLGLATPAIQVNGYTDRIKESANLKSQRAETVANYLVSKGNPANTYAVG 326
 OY 147 FGEERPIAFGTNEAMSQNRRA 168
 DB 327 YGEANPYTGATCDKV--KGRKA 346
 RESULT 35
 AAB44589
 ID AAB44589 standard; Protein: 369 AA.
 AC AAB44589;
 AC
 DT 08-FEB-2001 (first entry)
 XX
 DE Virulence gene protein #69.

XX
 KW Virulence gene: antibacterial; vaccine; bacterial infection;
 KW septicaemia; bronchopneumonia; rhinitis; wound infection.
 XX
 OS Actinobacillus pleuropneumoniae.
 XX
 FN WO200061724-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US09218.
 XX
 PR 09-APR-1999; 99US-0128689.
 PR 10-SEP-1999; 99US-0153453.
 XX
 PA (PHAA) PHARMACIA & UPJOHN INC.
 XX
 PI Lowery DE, Fuller TE, Kennedy MJ;
 PI
 XX
 DR WPI: 2000-647422/62.
 DR N-PSDB: AAC79664.
 XX
 PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence
 PT genes, useful as a live attenuated vaccine against bacterial infections
 XX
 PS Claim 39; Pages 308-309; 322pp; English.
 XX
 CC The family Pasteurellaceae encompasses several pathogens that infect a
 CC wide variety of animals. The present invention relates to virulence genes
 CC from Pasteurellaceae. The present sequence is a protein encoded by one
 CC such virulence gene. The virulence genes of the present invention may be
 CC mutated in order to produce an inactive gene. The inactive virulence gene
 CC may in turn be used to produce a vaccine, which is useful for treating
 CC bacterial infections such as septicaemias, bronchopneumonias, rhinitis and
 CC wound infections.
 CC
 SQ Sequence 369 AA:
 Query Match 19.6%; Score 169; DB 21; Length 369;
 Best Local Similarity 35.9%; Pred. No. 8, 5e-10;
 Matches 47; Conservative 16; Mismatches 58; Indels 10; Gaps 4;
 OY 34 APNAPGTYGVYIT---GVAPLVNDDETVKALSKLPISLYVDFDSDEIKPOAAALIDE 89
 DB 210 APDIHSTVAGLSYRFGQAVAPVEPEVYTKNFA--FSSDYLFDFGKSSLPAAATALD- 266
 OY 90 QAOFLITN--QTARVLYAGHTDERSREYMSIGERRAAVAVRYTLGKGINQASVEIIS 146
 DB 267 AANTEIANLGLATPAIQVNGYTDRIKESANLKSQRAETVANYLVSKGNPANTYAVG 326
 OY 147 FGEERPIAFGT 157
 DB 327 YGEANPYTGAT 337
 RESULT 36
 AAR66294
 ID AAR66294 standard; Protein: 359 AA.
 AC AAR66294;
 AC
 DT 09-AUG-1995 (first entry)
 XX
 DE Non-typable Haemophilus influenza (NTHI) fimbria protein.
 XX
 KW Fimbria protein; vaccine; otitis media.
 XX
 OS Haemophilus influenza strain 1128.
 XX
 PA
 XX
 PI Key Location/Qualifiers
 PI Region 22..33
 FT /label= amino terminus

FT Peptide 234..249
 FT /label= Internal CNBR fragment
 PN W09426304-A.
 XX 24-NOV-1994.
 PD
 XX
 PF 12-MAY-1994; 94WO-US05477.
 XX
 PR 18-MAY-1993; 93US-0065442.
 XX
 PA (OHIO-) OHIO STATE RES FOUND.
 XX
 PI Bakaletz LO, Kolattukudy PE, Sirakova T;
 DR WPI: 1995-006359/01.
 N-PSDB: AAQ78916.
 XX
 PT Vaccine comprising non-typable Haemophilus influenza fimbria
 PT protein - useful in studying, preventing or reducing the
 PT severity of otitis media, also fimbria protein and DNA.
 PS
 XX Disclosure; Fig 5; 45pp; English.
 CC The fimbria proteins from 15 randomly selected type b and non-
 CC typable clinical isolates of Haemophilus influenzae share common
 CC epitopes. Thus fimbria isolated from non-typable Haemophilus
 CC influenzae 1128 strain is a particularly suitable immunogen to
 CC protect against the different non-typable H₂ influenzae that cause
 CC otitis media. Fimbria protein is produced by culturing a transformed
 CC microbial host, pref. E.coli, Sporodoptera frugiperda or a mucosal
 CC pathogen. Fimbria protein (FP) produced by this process is claimed.
 CC The FP protein migrates in polyacrylamide gels to a posn. equiv. to
 CC a mol. wt. of 25.5 KD or 37.5 KD.
 CC
 SQ Sequence 359 AA;
 Query Match 19.2%; Score 165.5; DB 16; Length 359;
 Best Local Similarity 29.9%; Pred. No. 2e-09;
 Matches 47; Conservative 26; Mismatches 63; Indels 21; Gaps 4;
 QY 35 PNAPTGY-----TGVY---TGVAPLVNDDETVKALSKLPSTLYFDPDSDEKPPQA 83
 DB 199 PNAININPWIGICINAGISYRFGQGEAPVVAPEMVSKEFS-LNSDYTFATFKAMDKPPQA 257
 QY 84 AAILDEQAQFLITNOTARVAVAGTDERGSRREYNSLGERRAVAVNNYLLGKGINQASVE 143
 DB 258 QATLDVSYGEISQVSKRKAVALGTNRIGSDAFNNKLSQERADSVANFVAKGVAADALS 317
 QY 144 IISFGERRPIAFGTNEE-----AMSONRRALS 171
 DB 318 ATGTGEANPYTGATCDQVKGKRALIACIAPDRRVEIA 354
 RESULT 37
 AA01571
 ID AA01571 standard; Protein; 228 AA.
 AC AA01571;
 XX
 DT 19-JUL-2002 (first entry)
 DE M catarhalis MCA100854 protein SEQ ID NO: 22.
 XX
 KM Moraxella; vaccine; respiratory tract infection; antiinflammatory;
 XX auditory; antibacterial; otitis media; sinusitis; pneumonia.
 OS Moraxella catarhalis.
 XX
 PN W0200218595-A2.
 XX
 PD 07-MAR-2002.
 XX

PF 28-AUG-2001; 2001WO-CA01221.
 XX
 PR 28-AUG-2000; 2000US-228294P.
 PR 28-AUG-2000; 2000US-228295P.
 PR 28-AUG-2000; 2000US-228296P.
 PR 29-AUG-2000; 2000US-228438P.
 PR 29-AUG-2000; 2000US-228439P.
 PR 29-AUG-2000; 2000US-228440P.
 PR 29-AUG-2000; 2000US-228441P.
 PR 29-AUG-2000; 2000US-228442P.
 PR 29-AUG-2000; 2000US-228443P.
 PR 29-AUG-2000; 2000US-228511P.
 PR 29-AUG-2000; 2000US-228512P.
 PR 29-AUG-2000; 2000US-228742P.
 PR 01-SEP-2000; 2000US-228773P.
 PR 01-SEP-2000; 2000US-229455P.
 PR 01-SEP-2000; 2000US-229474P.
 PR 01-SEP-2000; 2000US-229475P.
 PR 01-SEP-2000; 2000US-229478P.
 PR 05-SEP-2000; 2000US-229740P.
 PR 05-SEP-2000; 2000US-229740P.
 PR 05-SEP-2000; 2000US-229803P.
 PR 05-SEP-2000; 2000US-229805P.
 PR 05-SEP-2000; 2000US-229806P.
 PR 05-SEP-2000; 2000US-229809P.
 PR 06-SEP-2000; 2000US-229811P.
 PR 06-SEP-2000; 2000US-230214P.
 PR 06-SEP-2000; 2000US-230250P.
 PR 06-SEP-2000; 2000US-230252P.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 PI
 DR Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;
 N-PSDB: AAL46503.
 XX
 DR WPI: 2002-401721/43.
 N-PSDB: AAL46503.
 XX
 PT Moraxella polypeptide and polynucleotides useful as vaccine for
 PT immunizing a host e.g. humans against disease e.g. otitis media,
 PT pneumonia, caused by infection of the bacteria
 XX
 Claim 28; Fig 21; 277pp; English.
 CC The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarhalis. These can be used to produce
 CC vaccines which protect against M. catarhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a protein of the invention.
 CC
 SQ Sequence 228 AA;
 Query Match 19.2%; Score 165; DB 23; Length 228;
 Best Local Similarity 30.7%; Pred. No. 1.2e-09;
 Matches 42; Conservative 22; Mismatches 71; Indels 2; Gaps 1;
 QY 33 VAPNAPTGYTGVYITGVAPLVNDDETVKALSKLPSTLYFDPDSDEKPPQAAILDEQAQ 92
 DB 68 VADSIETARVALG--DIVEENMDLINALNTQIINFALDSTETPOENKKEIDLAEE 125
 QY 93 FLITNOTARVAVAGTDERGSRREYNSLGERRAVAVNNYLLGKGINQASVEIISFGGERP 152
 DB 126 KIKAVPEETTLRIIGTDTQGTHEYNOQDLESRAAVALVEIYLSKQVAERLNTQASFPDY 185
 QY 153 IAFGTNEAMSONRRAL 169
 DB 186 VASNATEGGRFQNRRIE 202
 RESULT 38
 AA050718
 ID AA050718 standard; Protein; 344 AA.
 AC AA050718;
 XX

```

XX 18-APR-2002 (first entry)
XX Haemophilus paragallinarum strain 2403 haemagglutinin.
XX DE Haemagglutinin; HAgA; antigen; vaccine; immunisation; coryza;
XX KW chicken.
XX OS Haemophilus paragallinarum.
XX
XX Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /Label= Signal_peptide
XX FT 22..344
XX FT Protein /Label= Mature_protein
XX
XX WO200204485-A1.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-AU00822.
XX
XX 07-JUL-2000; 2000AU-0008652.
XX
XX (UYOU ) UNIV QUEENSLAND.
XX
XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX WPI; 2002-154917/20.
XX N-PSDB; ABA91419.
XX
XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX PT useful as a vaccine for immunising chickens against coryza caused by
XX PT the species -
XX
XX Claim 2; Fig 4; 67pp; English.
XX
XX The present sequence is that of the HAgA haemagglutinin of
XX CC Haemophilus paragallinarum strain 2403 (serovar A), the causative
XX CC agent of infectious coryza in chickens. The invention provides
XX CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX CC A, B and C) of H. paragallinarum, the causative agent of infectious
XX CC coryza of chickens. The polypeptides are useful in vaccines for
XX CC immunisation against infectious coryza, as are the nucleic acids
XX CC when expressed in attenuated bacteria, especially Salmonella or
XX CC Mycoplasma (claimed). The recombinant polypeptide is preferably
XX CC derivative, that is capable of eliciting an immune response,
XX CC providing protection against one or more strains of H.
XX CC paragallinarum in chickens. Also claimed are methods of using the
XX CC haemagglutinin polypeptides and nucleic acids for detection and
XX CC diagnosis of infectious coryza in chickens.
XX
XX Sequence 344 AA:
XX
XX Query Match 18.8%; Score 161.5; DB 23; Length 344;
XX Best Local Similarity 30.9%; Pred. No. 5.1e-09;
XX Matches 51; Conservative 25; Mismatches 68; Indels 21; Gaps 5;
XX
XX 22 GCANKSTSOVWVAPNAPGTGYIT--GVAPLYDNDETYKALASLPISLVFDPDSDEI 79
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 181 GREKDGSRVDYTPSIGVTAGLSYRFGQSAVPEPKVAKTFA--LNSDVTAFGKANL 238
XX
XX 80 KPOAAALID-----EQAQLTTNQTARVLVAGHTDERGSRREYNMISLGERRAVAVRNYLIGK 135
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 239 REPAQNVLDIGIYETAOL-----KSYOVVAGYTDIGISEANLKSORRADTVANYLYSK 294
XX
XX 136 GINQASVEIISFGEERPIAFGTFNE-----EAMSONRRAEIS 171
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 295 GVAQAEVISTGYGEANPVTGACQDVTYKGRKALLIACLAADRYREIS 339
XX

```

```

RESULT 39
AAM50721
ID AAM50721 standard; Protein; 344 AA.
XX
XX AAM50721;
XX
XX 18-APR-2002 (first entry)
XX Haemophilus paragallinarum strain 0222 haemagglutinin.
XX DE Haemagglutinin; HAgA; antigen; vaccine; immunisation; coryza;
XX KW chicken.
XX OS Haemophilus paragallinarum.
XX
XX Key Location/Qualifiers
XX FT Peptide 1..21
XX FT /Label= Signal_peptide
XX FT 22..344
XX FT Protein /Label= Mature_protein
XX
XX WO200204485-A1.
XX
XX 17-JAN-2002.
XX
XX 06-JUL-2001; 2001WO-AU00822.
XX
XX 07-JUL-2000; 2000AU-0008652.
XX
XX (UYOU ) UNIV QUEENSLAND.
XX
XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX WPI; 2002-154917/20.
XX N-PSDB; ABA91422.
XX
XX New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX PT useful as a vaccine for immunising chickens against coryza caused by
XX PT the species -
XX
XX Claim 2; Fig 4; 67pp; English.
XX
XX The present sequence is that of the HAgA haemagglutinin of
XX CC Haemophilus paragallinarum strain 0222 (serovar B), the causative
XX CC agent of infectious coryza in chickens. The invention provides
XX CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX CC A, B and C) of H. paragallinarum, the causative agent of infectious
XX CC coryza of chickens. The polypeptides are useful in vaccines for
XX CC immunisation against infectious coryza, as are the nucleic acids
XX CC when expressed in attenuated bacteria, especially Salmonella or
XX CC Mycoplasma (claimed). The recombinant polypeptide is preferably
XX CC derivative, that is capable of eliciting an immune response,
XX CC providing protection against one or more strains of H.
XX CC paragallinarum in chickens. Also claimed are methods of using the
XX CC haemagglutinin polypeptides and nucleic acids for detection and
XX CC diagnosis of infectious coryza in chickens.
XX
XX Sequence 344 AA:
XX
XX Query Match 18.8%; Score 161.5; DB 23; Length 344;
XX Best Local Similarity 30.9%; Pred. No. 5.1e-09;
XX Matches 51; Conservative 25; Mismatches 68; Indels 21; Gaps 5;
XX
XX 22 GCANKSTSOVWVAPNAPGTGYIT--GVAPLYDNDETYKALASLPISLVFDPDSDEI 79
XX | | | | | | | | | | | | | | | | | | | | | |
XX 181 GREKDGSRVDYTPSIGVTAGLSYRFGQSAVPEPKVAKTFA--LNSDVTAFGKANL 238
XX
XX 80 KPOAAALID-----EQAQLTTNQTARVLVAGHTDERGSRREYNMISLGERRAVAVRNYLIGK 135
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 239 REPAQNVLDIGIYETAOL-----KSYOVVAGYTDIGISEANLKSORRADTVANYLYSK 294
XX

```


22 GCANKSTSQVMVAPNAPTGYTYT--GVALVDNDETVKALASKLSLVYFDSDSEI 79
| | : | : : : | : | : |
181 GRVERDGSRYVTPTSPGTSVTEACI SVDECOGK DITTFE... : :

[illegible]

```
Search completed: July 6, 2003, 14:04:28
Job time : 64 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:14:15 ; Search time 16 Seconds

(without alignments)
445.871 Million cell updates/sec

Title: US-09-674-779B-2

Perfect score: 172
Sequence: 1 MMLHIQIAAAALSVLFM.....IAFGTNEAWSONRRRLSY 172Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	7.6	176	1	PAL_LEGPN
2	8	4.7	491	1	PBP_BACSU
3	8	4.7	829	1	E77A_DROME
4	7	4.1	37	1	ANP3_PSEAM
5	7	4.1	40	1	ANP8_MYOAE
6	7	4.1	82	1	ANPA_PSEAM
7	7	4.1	120	1	CU19_ARADI
8	7	4.1	127	1	CU26_ARADI
9	7	4.1	145	1	AZUP_PARPV
10	7	4.1	173	1	PAL_ECOLI
11	7	4.1	196	1	REMS_XENLA
12	7	4.1	200	1	RBMS_CHICK
13	7	4.1	203	1	CPPL_MOUSE
14	7	4.1	213	1	RLI4_RAT
15	7	4.1	251	1	COB1_MYCTU
16	7	4.1	261	1	GSHI_MOUSE
17	7	4.1	262	1	MYRK_METKA
18	7	4.1	266	1	MTRK_METTH
19	7	4.1	266	1	MTRK_METTH
20	7	4.1	290	1	HXB8_HUMAN
21	7	4.1	291	1	PSBO_CHLRE
22	7	4.1	294	1	RBSK_BACHD
23	7	4.1	304	1	GSH2_HUMAN
24	7	4.1	305	1	GSH2_MOUSE
25	7	4.1	322	1	GVPL_HAIME
26	7	4.1	322	1	SVPL_DROME
27	7	4.1	325	1	AB5B_MYCKA
28	7	4.1	330	1	GDC_BOVIN
29	7	4.1	338	1	RLAO_METUA
30	7	4.1	349	1	GUB_FIBSU
31	7	4.1	350	1	DKK3_HUMAN
32	7	4.1	354	1	KLF2_MOUSE
33	7	4.1	355	1	KLF2_HUMAN

34	7	4.1	367	1	BET3_MESAU	009029 mesocricetu
35	7	4.1	370	1	YCBU_BACSU	P42253 bacillus su
36	7	4.1	389	1	FL_ORYZA	O24175 oryza sativ
37	7	4.1	401	1	HB9_HUMAN	P50219 homo sapien
38	7	4.1	409	1	NH4M_ASCSU	P24880 ascaris suu
39	7	4.1	427	1	RL4_HUMAN	P36578 homo sapien
40	7	4.1	438	1	TUD4_AGRVI	O44472 agrobacteri
41	7	4.1	443	1	MURE_STRO	O43804 streptomyce
42	7	4.1	475	1	EYX2_MOUSE	P49749 mus musculu
43	7	4.1	476	1	EYX2_HUMAN	O03828 homo sapien
44	7	4.1	478	1	FXD3_HUMAN	O94955 homo sapien
45	7	4.1	481	1	KNIR_DROYI	O24753 drosophila
46	7	4.1	481	1	BHB3_HUMAN	O9c0j9 homo sapien
47	7	4.1	485	1	NAB1_MESAU	O35589 mesocricetu
48	7	4.1	485	1	NAB1_RAT	O62722 rattus norv
49	7	4.1	486	1	NAB1_HUMAN	O13506 homo sapien
50	7	4.1	486	1	NAB1_MOUSE	O61122 mus musculu
51	7	4.1	492	1	TISD_HUMAN	P47974 homo sapien
52	7	4.1	496	1	BAFL_KLUMA	P33293 kluveromyc
53	7	4.1	498	1	CAUR_PIG	P25117 sus scrofa
54	7	4.1	498	1	LYXK_ECOLI	P37677 escherichia
55	7	4.1	503	1	TGRI_HUMAN	P36897 homo sapien
56	7	4.1	517	1	VSM5_TRYBR	P21840 trypanosoma
57	7	4.1	530	1	ZIC2_MOUSE	O62520 mus musculu
58	7	4.1	532	1	ZIC2_HUMAN	O95409 homo sapien
59	7	4.1	537	1	MMSA_BOVIN	O07536 bos tauris
60	7	4.1	547	1	MCPC_SALTY	O02755 salmonella
61	7	4.1	551	1	MCPI_ECOLI	P02942 escherichia
62	7	4.1	553	1	MCPI_ECOLI	P07017 escherichia
63	7	4.1	553	1	MCPI_ECOLI	P02941 salmonella
64	7	4.1	622	1	SR68_CANFA	O00004 canis fami1
65	7	4.1	647	1	KNRL_DROME	P13054 drosophila
66	7	4.1	665	1	LI23_CAEEL	O00990 caenorhabdi
67	7	4.1	671	1	HMOC_DROME	P22810 drosophila
68	7	4.1	688	1	YHIO_YEAST	P38708 saccharomyc
69	7	4.1	741	1	ERF2_PICPI	P23637 pichia pinu
70	7	4.1	766	1	EVA_DROME	O05201 drosophila
71	7	4.1	810	1	REB1_YEAST	P21538 saccharomyc
72	7	4.1	883	1	E74B_DROME	P11536 drosophila
73	7	4.1	884	1	YAVK_CAEEL	P17851 caenorhabdi
74	7	4.1	912	1	KPCM_HUMAN	O15139 homo sapien
75	7	4.1	918	1	KPCM_MOUSE	O62101 mus musculu
76	7	4.1	988	1	OMB_DROME	O24432 drosophila
77	7	4.1	1047	1	CUGA_ECOLI	P38054 escherichia
78	7	4.1	1092	1	DP2L_METTH	O27579 methanobact
79	7	4.1	1225	1	CTD2_HUMAN	O9ub33 homo sapien
80	7	4.1	1385	1	NAC2_CHLRE	O91ame8 chlamydomon
81	7	4.1	2038	1	FSH_DROME	P13709 drosophila
82	7	4.1	2175	1	HMCU_DROME	P10180 drosophila
83	7	4.1	2799	1	EDD_HUMAN	O95071 homo sapien
84	7	3.5	33	1	ANP3_MYOSC	P04367 myoxocephal
85	6	3.5	45	1	ANP8_MYOAE	P04368 myoxocephal
86	6	3.5	45	1	ANP8_MYOAE	P04368 myoxocephal
87	6	3.5	52	1	RL4O_YEAST	P14796 saccharomyc
88	6	3.5	53	1	RL4O_CRYNE	P49039 cryptococcu
89	6	3.5	55	1	RL32_VIBCH	O9k437 vibrio chol
90	6	3.5	65	1	OMP_LOCOM	P80045 locustiana t
91	6	3.5	72	1	RL4O_TOBAC	P19379 nicotiana t
92	6	3.5	74	1	SR14_MACRA	O18881 macera radi
93	6	3.5	85	1	ANP4_PSEAM	P02734 pseudopleur
94	6	3.5	85	1	MRKD_RAT	P20468 rattus norv
95	6	3.5	87	1	YHCN_ECOLI	P46477 escherichia
96	6	3.5	90	1	DBHA_AERHY	O91496 aeromonas h
97	6	3.5	90	1	DBHA_VIBPR	P28080 vibrio prot
98	6	3.5	91	1	ANPX_PSEAM	P07835 pseudopleur
99	6	3.5	91	1	ANPX_PSEAM	P23699 pseudopleur
100	6	3.5	97	1	ANP_LTIME	P09031 limanda fer

ALIGNMENTS

RESULT 1

PAL_LEGPN
ID PAL_LEGPN STANDARD; PRT; 176 AA.
AC P26493;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Peptidoglycan-associated lipoprotein precursor (19 kDa surface antigen) (PPL).
GN PAL OR PPLA.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AI100 / Serogroup 1;
RX MEDLINE=92114778; PubMed=1766377;
RA Engleberg N.C., Howe D.C., Rogers J.E., Arroyo J., Eisenstein B.I.;
RT "Characterization of a Legionella pneumophila gene encoding a
RT lipoprotein antigen."
RL MOL. Microbiol. 5:2021-2029(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91310296; PubMed=1855972;
RA Ludwig B., Schmid A., Maire R., Hacker J.;
RT "Cloning, genetic analysis, and nucleotide sequence of a determinant
RT coding for a 19-kilodalton peptidoglycan-associated protein (Ppl) of
RT Legionella pneumophila."
RL Infect. Immun. 59:2515-2521(1991).
CC -1- FUNCTION: VERY STRONGLY ASSOCIATED WITH THE PEPTIDOGLYCAN.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -1- SIMILARITY: TO OTHER PAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X60543; CAA43033.1; -
DR PIR: A60337; A60337.
DR PIR: S16631; S16631.
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_OmpA; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE: PS01068; OMPA; 1.
KW Lipoprotein; Outer membrane; Antigen; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 176 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
FT LIPID 22 22 N-ACYL DIGLYCERIDE (PROBABLE).
FT DOMAIN 104 148 OMPA-LIKE.
SQ SEQUENCE 176 AA; 18911 MW; 7D9C3BECE6E21DB CRC64;
Query Match 176; Score 13; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. NO. 5,2e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative penicillin binding protein precursor.
GN PBP OR DACC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168.
RX MEDLINE=95227362; PubMed=7711903;
RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
RA Grandi G.;
RT "A putative new peptide synthase operon in Bacillus subtilis: partial
RT characterization."
RL Microbiology 141:645-648(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berrero M.G., Bessieres P., Bolojin A., Borchert S.,
RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kunano M.,
RA Kurita K., Lapins A., Lardinois S., Lauber J., Lazzeric V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Meliado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Portwolk S., Prescott A.M.,
RA Prescan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Welleneger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S13; ALSO KNOWN AS THE
CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 3 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z34883; CAA84366.1; -
DR EMBL: Z99113; CAB13718.1; -
DR MEROPS: S13.0PW; -
DR Subtilist: BG10969; pdp.
DR InterPro: IPR000667; Peptidase_S13.
DR Pfam: PF02113; Peptidase_S13; 1.
DR PRINTS: PR00922; DADCBPTASE3.
DR TIGRFAMs: TIGR00666; PBP4; 1.
KW Peptidoglycan synthesis; Cell division; Cell wall; Hydrolase; Signal;
KW Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 491 PUTATIVE PENICILLIN BINDING PROTEIN.

FT ACT_SITE 81 81 ACYLATED BY PENICILLIN (FORMS PART OF
 FT SEQUENCE 491 AA; 52891 MW; E442E5A227BD080 CRC64;
 THE ACTIVE-SITE CLEFT) (BY SIMILARITY).
 Query Match 4.7%; Score 8; DB 1; Length 491;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AAAALSVL 17
 DB 88 AAAALSVL 95

RESULT 3
 E74A_DROME STANDARD; PRT; 829 AA.
 ID E74A_DROME
 AC P20105;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ecdysone-induced protein 74EF isoform A (ETS-related protein E74A).
 GN Eir74EF OR E74.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 [1]
 RN SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RP MEDLINE=90199900; PubMed=2107982;
 RA Butts K.C., Thumel C.S., Jones C.W., Karim F.D., Hogness D.S.;
 RT "The Drosophila 74EF early puff contains E74, a complex ecdysone-
 inducible gene that encodes two ets-related proteins.";
 RL Cell 61:85-99(1990).
 [2]
 RN DEVELOPMENTAL STAGE.
 RP MEDLINE=94038699; PubMed=8223281;
 RA Huet F., Ruiz C., Richards G.;
 RT "Puffs and PCR: the in vivo dynamics of early gene expression during
 ecdysone responses in Drosophila.";
 RL Development 118:613-627(1993).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A/E74A (shown here) and B/E74B
 (AC P11536); are produced by alternative splicing.
 CC -1- DEVELOPMENTAL STAGE: In mid instar larvae salivary glands levels
 are low during puff stage 1 and increase during puff stage 2 to
 become the predominant form in stage 3. Levels reach maximum in
 late larvae during puff stages 8-10, decreasing abruptly at stage
 11. This expression pattern is also seen in Malpighian tubules and
 wing disk. Levels at puff stage 11 are appreciable in the gut and
 fat body. Transcripts are detected again in salivary glands from
 puff stages 12-14 and 17-21.
 CC -1- INDUCTION: THE EXPRESSION OF THIS PROTEIN IS DEVELOPMENTALLY
 REGULATED AND IS CORRELATED WITH THE 20-OH-ECDYSONE INDUCED
 ACTIVITY OF PUFF 74EF.
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL: M37082; AAA28493.1;
 DR PIR: A34692; A34692.
 DR HSSP: P14921; 25ST.
 DR TRANSFAC: T00208;
 DR FlyBase: FBgn000567; Eir74EF.
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF-ETS.
 DR Pfam: PF00178; Ets; 1.

DR PRINTS: PR00454; ETSDOMAIN.
 DR SMART: SM00413; ETS; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS0061; ETS_DOMAIN_3; 1.
 KW Nuclear protein; transcription regulation; DNA-binding;
 KW Alternative splicing.
 FT DNA_BIND 733 815
 FT DOMAIN 126 129
 FT DOMAIN 187 190
 FT DOMAIN 191 201
 FT DOMAIN 208 219
 FT DOMAIN 227 232
 FT DOMAIN 318 324
 FT DOMAIN 352 361
 FT DOMAIN 380 383
 FT DOMAIN 432 435
 FT DOMAIN 456 470
 FT DOMAIN 560 576
 FT DOMAIN 585 588
 FT DOMAIN 597 607
 FT DOMAIN 641 649
 FT DOMAIN 653 657
 FT DOMAIN 658 663
 FT DOMAIN 667 679
 FT VARIANT 813 813
 H -> Q.
 SO SEQUENCE 829 AA; 87147 MW; 6E3CF7CF0A1D7F5D CRC64;

QY 7 IAAAAAL 14
 DB 195 IAAAAAL 202

RESULT 4
 ANP3_PSEAM STANDARD; PRT; 37 AA.
 ID ANP3_PSEAM
 AC P02733;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antifreeze peptide 3.
 OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
 americanus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Pleuronectidae; Pleuronectidae; Pseudopleuronectes.
 NX NCBI_TaxID=8265;
 RN [1]
 RN SEQUENCE.
 RP MEDLINE=78060969; PubMed=588591;
 RA Devries A.L., Lin Y.;
 RT "Structure of a peptide antifreeze and mechanism of adsorption to
 ice.";
 RL Biochim. Biophys. Acta 495:388-392(1977).
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
 ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
 DR PIR: A03192; FDFL3W.
 DR InterPro: IPR000104; Antifreeze_1.
 DR PRINTS: PR00308; ANTIFREEZE1.
 KW Antifreeze protein; Repeat; Multigene family.
 SO SEQUENCE 37 AA; 3144 MW; 46AA951A962DECA9 CRC64;

QY 8 AAAAAL 14

Query Match 4.1%; Score 7; DB 1; Length 37;
 Best Local Similarity 100.0%; Pred. No. 6.2;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 6 AAAAAAL 12

RESULT 5

ANP8_MYOE STANDBO: PRT: 40 AA.
 ID ANP8_MYOE
 AC P20617;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE Antifreeze peptide GS-8.
 OS Myoxocephalus aeneus (Grubby sculpin).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
 OC Cottidae; Cottidae; Myoxocephalus.
 OC NCBI_Taxid=8096;
 RX [1]
 RN [1]
 RA Chakrabarty A., Hew C.-L., Shears M., Fletcher G.;
 RT "Primary structures of the alanine-rich antifreeze polypeptides from
 RT grubby sculpin, Myoxocephalus aeneus.";
 RL Can. J. Zool. 66:403-408(1988).
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
 CC PIR: S07046; FDF18G.
 DR Antifreeze protein. Repeat. BLOCKED.
 KW MOD_RES 1
 FT MOD_RES 1
 SQ SEQUENCE 40 AA; 3579 MM; 32F50E243C2AD11 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 40;
 Best Local Similarity 100.0%; Pged. No. 6; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0;

OY 8 AAAAAAL 14
 Db 14 AAAAAAL 20

RESULT 6

ANP8_PSEAM STANDBO: PRT: 82 AA.
 ID ANP8_PSEAM
 AC P04002;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antifreeze protein A/B precursor.
 OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
 OS americanus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Pleuronectidae; Pleuronectidae; Pseudopleuronectes.
 OC NCBI_Taxid=8265;
 RX [1]
 RN [1]
 RA SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=82197490; PubMed=6952188;
 RA Davies P.L., Roach A.H., Hew C.-L.;
 RT "DNA sequence coding for an antifreeze protein precursor from winter
 RT flounder.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
 RN [2]
 RN [2]
 RA SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=88259236; PubMed=3133486;
 RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
 RT "Differential amplification of antifreeze protein genes in the
 RT pleuronectidae.";
 RL J. Mol. Evol. 27:29-35(1988).
 RN [3]
 RN [3]
 RA SEQUENCE FROM N.A. (PROTEIN B).
 RX MEDLINE=84264559; PubMed=6086629;

RA Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
 RT "Antifreeze protein genes of the winter flounder.";
 RL J. Biol. Chem. 259:9241-9247(1984).
 RN [4]
 RN [4]
 RA SEQUENCE FROM N.A.
 RX MEDLINE=92209995; PubMed=1555765;
 RA Davies P.L.;
 RT "Conservation of antifreeze protein-encoding genes in tandem
 RT repeats.";
 RL Gene 112:163-170(1992).
 RN [5]
 RN [5]
 RA 3D-STRUCTURE MODELING OF 45-81.
 RX MEDLINE=92148833; PubMed=1738160;
 RA Chou K.-C.;
 RT "Energy-optimized structure of antifreeze protein and its binding
 RT mechanism.";
 RL J. Mol. Biol. 223:509-517(1992).
 RN [6]
 RN [6]
 RA X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
 RX MEDLINE=95281060; PubMed=7760940;
 RA Siebert F., Yang D.S.C.;
 RT "Ice-binding structure and mechanism of an antifreeze protein from
 RT winter flounder.";
 RL Nature 375:427-431(1995).
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE I AFP ARE
 CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: L00138; AAB59964.1;
 DR EMBL: L29178; AAB59964.1; JOINED.
 DR EMBL: M62414; AAA49469.1;
 DR EMBL: X07506; CAA30389.1;
 DR EMBL: M62416; AAA49471.1;
 DR EMBL: M62417; AAA49472.1;
 DR PIR: A03194; FDF18G.
 DR PIR: A05161; A05161.
 DR PIR: S02326; S02326.
 DR PIR: JS0704; JS0704.
 DR PDB: 1ATF; 15-OCT-94.
 DR PDB: 1WEA; 03-JUN-95.
 DR PDB: 1WEF; 03-JUN-95.
 DR InterPro: IPR000104; Antifreeze_1.
 DR PRINTS: PR00308; ANTIFREEZE1.
 KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
 FT SIGNAL 1
 FT PROPEP 22
 FT CHAIN 45
 FT VARIANT 36
 FT VARIANT 70
 FT CONFLICT 24
 FT HELIX 46
 SQ SEQUENCE 82 AA; 7711 MM; C2AE7B74C0D46CC1 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 82;
 Best Local Similarity 100.0%; Pged. No. 12;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAL 14
 Db 50 AAAAAAL 56

RESULT 7

CUI9_ARADI
ID CUI9_ARADI STANDARD; PRT; 120 AA.
AC P80515;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adult-specific rigid cuticular protein 11.9 (ACP 11.9).
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
OX NCBI_TaxID=45920;
RN [1]
RP TISSUE=Cuticle;
RX MEDLINE=9716616; PubMed=9014336;
RA Norup T., Berg T., Stenholm H., Andersen S.O., Hoejrup P.;
RT Purification and characterization of five cuticular proteins from
RL the spider Araneus diadematus.;
CC -1- FUNCTION: COMPONENT OF THE RIGID CUTICLE OF THE SPIDER.
CC -1- MASS SPECTROMETRY: MW=11872.1; METHOD=Electrospray.
DR InterPro: IPR000618; Insect_cuticle.
DR Pfam: PF00379; Chitin_bind_4; 1.
DR PRINTS: PR00947; CUTICLE.
DR PROSITE: PS00233; CUTICLE; 1.
KW Structural protein; Cuticle.
FT DOMAIN 73
SQ SEQUENCE 120 AA; 11871 MW; 803E577C61999DC2 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 73 AAAAAA 79

RESULT 8
CUI9_ARADI
ID CUI9_ARADI STANDARD; PRT; 127 AA.
AC P80517;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adult-specific rigid cuticular protein 12.6 (ACP 12.6).
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
OX NCBI_TaxID=45920;
RN [1]
RP TISSUE=Cuticle;
RX MEDLINE=9716616; PubMed=9014336;
RA Norup T., Berg T., Stenholm H., Andersen S.O., Hoejrup P.;
RT Purification and characterization of five cuticular proteins from
RL the spider Araneus diadematus.;
CC -1- FUNCTION: COMPONENT OF THE RIGID CUTICLE OF THE SPIDER.
CC -1- MASS SPECTROMETRY: MW=12587.5; METHOD=Electrospray.
DR InterPro: IPR000618; Insect_cuticle.
DR Pfam: PF00379; Chitin_bind_4; 1.
DR PRINTS: PR00947; CUTICLE.
DR PROSITE: PS00233; CUTICLE; 1.
KW Structural protein; Cuticle.
FT DOMAIN 73
SQ SEQUENCE 127 AA; 12589 MW; 9BFBCEB2B6AAB9 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 73 AAAAAA 79

RESULT 9
AZUP_PARP_N
ID AZUP_PARP_N STANDARD; PRT; 145 AA.
AC P80401;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pseudoazurin precursor.
GN PADS.
OS Paracoccus pantotrophus (Rhiosphaera pantotropha).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group.
OC Paracoccus.
OX NCBI_TaxID=82367;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-32.
RX STRAIN=ATCC 35512 / LMD 82.5;
RX MEDLINE=97184655; PubMed=9032456;
RA Leung Y.-C., Chan C., Reader J.S., Willis A.C., van Spanning R.J.M.,
RA Ferguson S.J., Radford S.E.,
RT "The pseudoazurin gene from Rhiosphaera pantotropha: analysis of
RT upstream putative regulatory sequences and overexpression in
RT Escherichia coli.";
RL Biochem. J. 321:699-705(1997).
RN [2]
RP SEQUENCE OF 23-145.
RX STRAIN=ATCC 35512 / LMD 82.5;
RX MEDLINE=95289994; PubMed=7772045;
RA Chan C., Willis A.C., Robinson C.V., Aplin R.T., Radford S.E.,
RA Ferguson S.J.;
RT "The complete amino acid sequence confirms the presence of
RT pseudoazurin in Rhiosphaera pantotropha.";
RL Biochem. J. 308:585-590(1995).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RA Williams P.A.;
RL Submitted (FEB-1997) to the PDB data bank.
CC -1- FUNCTION: THIS SOLUBLE ELECTRON TRANSFER COPPER PROTEIN IS
CC REQUIRED FOR THE INACTIVATION OF COPPER-CONTAINING NITRITE
CC REDUCTASE IN THE PRESENCE OF OXYGEN.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: Z73141; CAA97485.1; -
CC EMBL: Z70033; CAA93848.1; -
CC PDB: 1ADM; 15-MAY-97.
CC InterPro: IPR000923; BlueCu_1.
CC InterPro: IPR001235; Copper_blue.
CC Pfam: PF00127; copper_bind; 1.
CC PRINTS: PR00156; COPPERBLUE.
CC ProDom: PD001235; Copper_blue; 1.
CC PROSITE: PS00196; COPPER_BLUE; 1.
KW Copper; Electron transport; Signal; Periplasmic; 3D-structure.
FT SIGNAL 1
FT CHAIN 23
FT DOMAIN 27 115
FT METAL 62
FT METAL 100
FT METAL 103
FT METAL 103
PSEUDOAZURIN.
COPPER (BY SIMILARITY).
COPPER (BY SIMILARITY).
COPPER (BY SIMILARITY).
COPPER (BY SIMILARITY).

FT METAL 108 108 COPPER (BY SIMILARITY).
 SQ SEQUENCE 145 AA; 15446 MW; 725ECB5929EC3831 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 145;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
 7 AAAAAAL 13

RESULT 10
 ID PAL_ECOLI STANDARD; PRT; 173 AA.
 AC P07176;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL OR EXCC OR B0741 OR Z0909 OR ECS0776.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=871133578; PubMed=3545827;
 RA Chen R., Henning U.;
 RT "Nucleotide sequence of the gene for the peptidoglycan-associated
 RT lipoprotein of Escherichia coli K12.";
 RL Eur. J. Biochem. 163:73-77(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=92244043; PubMed=1574003;
 RA Lazzaroni J.-C., Portallier R.;
 RT "The excC gene of Escherichia coli K-12 required for cell envelope
 RT integrity encodes the peptidoglycan-associated lipoprotein (Pal).";
 RL Mol. Microbiol. 6:735-742(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horikuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsuda E., Nakayama K., Murata T., Tanaka M., Toke T.,
 RA Iida T., Takami H., Honda T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP SEQUENCE OF 1-30 FROM N.A.
 RC STRAIN-JM105;
 RX MEDLINE=90078104; PubMed=2687247;
 RA Levengood S.K., Webster R.E.;
 RT "Nucleotide sequences of the tolA and tolB genes and localization of
 RT their products, components of a multistep translocation system in
 RT Escherichia coli.";
 RL J. Bacteriol. 171:6600-6609(1989).
 RN [8]
 RP CRYSTALLIZATION.
 RX MEDLINE=21111468; PubMed=11173492;
 RA Abargel C., Walburger A., Chenivesse S., Lazdunski C.;
 RT "Crystallization and preliminary crystallographic study of the
 RT peptidoglycan-associated lipoprotein from Escherichia coli.";
 RL Acta Crystallogr. D 57:317-319(2001).
 CC -1- FUNCTION: Thought to play a role in bacterial envelope integrity.
 CC -1- Very strongly associated with the peptidoglycan.
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.
 CC -1- SIMILARITY: TO OTHER PAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X05123; CAA28771.1; -;
 CC EMBL: X65796; CAA4673.1; -;
 CC EMBL: AE000177; AAC73835.1; -;
 CC EMBL: D90713; BAA35407.1; -;
 CC EMBL: AE005252; AAG55077.1; -;
 CC EMBL: AP002553; BAB34199.1; -;
 CC EMBL: M28232; -; NOT_ANNOTATED_CDS.
 CC PIR: A27534; LPECPG.
 CC PIR: S20547; S20547.
 CC EcoGene: EG10684; pal.
 CC InterPro: IPR001145; Bac_OmpA.
 CC Pfam: PF00691; OmpA; 1.
 CC PRINTS: PR01021; OMPADOMAIN.
 CC PRODOM: PD000930; Bac_OmpA; 1.
 CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC PROSITE: PS01068; OMPA; 1.
 CC Outer membrane; Signal; Lipoprotein; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 173 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
 FT LIPID 22 N-ACYL DIGLYCERIDE.
 FT DOMAIN 105 149 OMPA-LIKE.
 SQ SEQUENCE 173 AA; 18824 MW; 449F9959C0274430 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 173;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SLGERRA 125

Db 120 SLGERRA 126

RESULT 11

ID RBMS_XENLA STANDARD; PRT; 196 AA.

AC Q9YGP5;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE RNA-binding protein with multiple splicing homolog (RBP-MS) (Heart,

DE RRM Expressed Sequence) (Hermes).

GN RBPM5 OR HERMES.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodidae; Xenopus.

OX NCBI_TaxID=8335;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=99195799; PubMed=10096065;

RA Gerber W.V., Yatskevich T.A., Antin P.B., Correia K.M., Conlon R.A.,

RA Krieg P.A.;

RT "The RNA-binding protein gene, hermes, is expressed at high levels in

RT the developing heart."

RL Mech. Dev. 80:77-86(1999).

CC -1- FUNCTION: MAY BIND RNA.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING HEART, PRONEPHROS,

CC RETINA AND EPIPHYSIS. IN ADULT, HIGH EXPRESSION IN HEART, MODERATE

CC IN KIDNEY, UNDETECTABLE IN LIVER, LONG AND SKELETAL MUSCLE.

CC -1- DEVELOPMENTAL STAGE: MRNA FIRST DETECTED IN THE TAILBUD EMBRYO

CC (STAGE 26) IN THE PAIRED HEART PRIMORDIA AND IN THE CONDENSING

CC EPITHELIUM THAT WILL FORM THE PRONEPHROS. AT THE LATE TAILBUD

CC STAGE (STAGE 34) IN THE DEVELOPING RETINA AND EPIPHYSIS. AS

CC DEVELOPMENT PROCEEDS, DETECTED THROUGH THE ENTIRE LENGTH OF THE

CC HEART TUBE, IN THE MUSCULAR TISSUE OF THE OUTFLOW TRACT, AND IN

CC THE DUCT EPITHELIUM OF THE PRONEPHROS. DURING LATER DEVELOPMENT,

CC MRNA FOUND IN ALL SUBREGIONS OF THE HEART, IN THE GLOMUS, TUBULES

CC AND DUCT OF THE PRONEPHROS, IN THE RETINAL GANGLION CELL LAYER

CC (GCL) AND IN THE EPIPHYSIS.

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AF107889; AAD16971.1; -

CC HSSP: P09012; 201A.

CC InterPro: IPR000504; RNA_rec_mot.

CC Pfam: PF00076; rrm; 1.

CC SMART: SM00360; RRM; 1.

CC PROSITE: PS50102; RRM; 1.

CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.

CC RNA-binding.

CC FT DOMAIN 20 97 RNA-BINDING (RRM).

CC FT DOMAIN 11 14 POLY-ASN.

CC FT DOMAIN 167 172 POLY-ALA.

CC SEQUENCE 196 AA; 21675 MW; 7964198C1122A3D3 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 196;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 AAAAAL 14

Db 167 AAAAAL 173

RESULT 12

ID RBMS_CHICK STANDARD; PRT; 200 AA.

AC Q9W611;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE RNA-binding protein with multiple splicing homolog (RBP-MS) (Heart,

DE RRM Expressed Sequence) (Hermes).

GN RBPM5 OR HERMES.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=99195799; PubMed=10096065;

RA Gerber W.V., Yatskevich T.A., Antin P.B., Correia K.M., Conlon R.A.,

RA Krieg P.A.;

RT "The RNA-binding protein gene, hermes, is expressed at high levels in

RT the developing heart."

RL Mech. Dev. 80:77-86(1999).

CC -1- FUNCTION: MAY BIND RNA.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING HEART.

CC -1- DEVELOPMENTAL STAGE: MRNA ALREADY DETECTED AT STAGE 7-8 IN THE

CC CARDIOGENIC MESODERM, AND BECOME ALMOST UNDETECTABLE IN THE OUTER

CC CURVATURE OF THE VENTRICULAR REGION WHEREAS REMAINING HIGH IN THE

CC DEVELOPING ATRIAL REGIONS.

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AF129933; AAD30273.1; -

CC InterPro: IPR000504; RNA_rec_mot.

CC Pfam: PF00076; rrm; 1.

CC SMART: SM00360; RRM; 1.

CC PROSITE: PS50102; RRM; 1.

CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.

CC RNA-binding.

CC FT DOMAIN 22 99 RNA-BINDING (RRM).

CC FT DOMAIN 169 176 POLY-ALA.

CC SEQUENCE 200 AA; 21856 MW; 4093B3C780BDC1DF CRC64;

Query Match 4.1%; Score 7; DB 1; Length 200;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 AAAAAL 14

Db 171 AAAAAL 177

RESULT 13

ID CTFL_MOUSE

AC Q60753;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Cardiotrophin-1 (CT-1).

GN CTFL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
MEDLINE=95166785; PubMed=7862649;
RA Penica D., King K.L., Shaw K.J., Luis E., Rullamas J., Luch S.-M.,
RA Darbonne W.C., Knutson D.S., Yen R., Chien K.R., Baker J.B.,
RA Wood M.I.;
RT "Expression cloning of cardiostrophin 1, a cytokine that induces
RT cardiac myocyte hypertrophy." 92.1142-1146(1995).
RL Proc. Natl. Acad. Sci. U.S.A. 92.1142-1146(1995).
CC -1- FUNCTION: INDUCES CARDIAC MYOCYTE HYPERTROPHY IN VITRO. BINDS TO
CC AND ACTIVATES THE LEUKEMIA INHIBITORY FACTOR RECEPTOR (LIF
CC RECEPTOR)/GP 130 RECEPTOR COMPLEX.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,
CC LIVER, LUNG AND KIDNEY. LOWER LEVELS IN TESTIS AND BRAIN. NO
CC EXPRESSION IN SPLEEN.
CC -1- SIMILARITY: BELONGS TO THE IL-6 SUPERFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: U18366; AAC52173.1; -
DR MGD: MGI:105115; Ctlf.
KW Cytokine.
SQ SEQUENCE 203 AA; 21509 MW; 8B3D414A0B3B232F CRC64;

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AAATSVL 17
Db 91 AAATSVL 97

RESULT 14
ID RL14_RAT STANDARD; PRT; 213 AA.
AC 063507;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L14.
DE RPL14.
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40 AND 119-141.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96222520; PubMed=8670222;
RA Chan Y.-L., Olivera J., Wool I.G.;
RT "The primary structure of rat ribosomal protein L14."
RL Biochem. Biophys. Res. Commun. 222:427-431(1996).
CC -1- SIMILARITY: BELONGS TO THE L14E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: X94242; CAA63926.1; -
DR InterPro: IPR000302; K0W_motif.

InterPro: IPR002784; Ribosomal_L14e.
DR Pfam: PF00467; K0W_1.
DR Pfam: PF01929; Ribosomal_L14e; 1.
KW Ribosomal protein; Repeat.
FT INTU_MET 0
FT DOMAIN 169 188 4 X 5 AA TANDEM REPEATS OF Q-K-A-[AS]-X.
FT REPEAT 169 173 1-1.
FT REPEAT 169 178 1-2.
FT REPEAT 174 183 1-3.
FT REPEAT 179 183 1-4.
FT REPEAT 184 188 1-4.
FT DOMAIN 192 197 2 X 3 AA TANDEM REPEATS OF K-G-Q.
FT REPEAT 192 194 2-1.
FT REPEAT 195 197 2-2.
SQ SEQUENCE 213 AA; 23207 MW; E14856D6070A1A44 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 213;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAA 13
Db 149 IAAAAA 155

RESULT 15
ID COBM_MYCTU STANDARD; PRT; 251 AA.
AC 010672;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Precorrin-4 C11-methyltransferase (EC 2.1.1.133) (Precorrin-3
DE methylase).
GN COBM OR RV2071C OR MT2131 OR MTCY49.10C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence." Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Debay R., Dodson R., Gwin M.L., Hart D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains." Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL C-11 IN PRECORRIN-4 TO FORM PRECORRIN-5.
CC -1- FUNCTION: METHYLTRANSFERASE THAT CATALYZES THE METHYLATION OF
CC C-11 IN PRECORRIN-4 TO FORM PRECORRIN-5.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-4 = S-
CC adenosyl-L-homocysteine + precorrin 5.
CC -1- PATHWAY: Cobalamin biosynthesis.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC AND CBIL/COBI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z73966; CAA98218.1; -
DR EMBL; AE007063; AAK46411.1; ALT_INT.
DR TIGR; MT2131; -
DR TubercuList; RV2071C; -
DR InterPro: IPR000878; Cor/Por_Mettransf.
DR InterPro: IPR003043; Uropor_Mettransf.
DR Pfam: PF00590; TP_methylase; 1.
DR PROSITE; PS00839; SUMP_1; 1.
DR PROSITE; PS00840; SUMP_2; 1.
KW Cobalam biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 251 AA; 26420 MW; 8D8F21ABFC91CD3 CRC64;

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 251;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAL 14
DB 114 AAAAAAL 120

RESULT 16
GSH1_MOUSE
ID GSH1_MOUSE STANDARD; PRT; 261 AA.
AC P31315;

DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein GSH-1.
GN GSH1 OR GSH-1.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=NIH Swiss;
RX MEDLINE=96172995; PubMed=8589431;
RA Valerius M.T., Li H., Stock J.L., Weinstein M., Kaur S., Singh G.,
RA Potter S.S.;
RT "Gsh-1: a novel murine homeobox gene expressed in the central nervous
RT system.";
RL Dev. Dyn. 203:337-351(1995).

RN [2]
RP SEQUENCE FROM N.A. AND FUNCTION.
RX MEDLINE=96181350; PubMed=8631293;
RA Li H., Zeitler P.S., Valerius M.T., Small K., Potter S.S.;
RT "Gsh-1, an orphan Hox gene, is required for normal pituitary
RT development.";
RL EMBO J. 15:714-724(1996).

RN [3]
RP SEQUENCE OF 146-205 FROM N.A.
RX MEDLINE=92073356; PubMed=1683707;
RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Potter S.S.;
RT "Identification of 10 murine homeobox genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).

CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA
CC SEQUENCE 5'-GC[TA][AC]ATTA[GA]-3'. ACTIVATES THE TRANSCRIPTION OF
CC -1- SUBCELLULAR LOCATION: Nuclear
CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U21224; AAA96814.1; -
DR PIR; A37290; A37290.
DR PIR; A38809; A38809.
DR HSSP; P14653; 1B72.
DR MGD; MGI:95842; Gsh1.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00711; HOMEBOX_2; 1.
KW Transcription regulation; Activator; Homeobox; DNA-binding;
KW Nuclear protein; Developmental protein.
FT DOMAIN 110 117
FT DNA_BIND 146 205
SQ SEQUENCE 261 AA; 27727 MW; BA9D6B94359E4FF3 CRC64;

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 261;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAL 14
DB 112 AAAAAAL 118

RESULT 17
PYRK_METKA
ID PYRK_METKA STANDARD; PRT; 262 AA.
AC P58887;

DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable dihydroorotate dehydrogenase electron transfer subunit.
GN PYRK OR MK0564.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus

OX NCBI_TaxID=2320;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Sheherbina O.V., Shakhova V.V., Belova G.I., Arayind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Mal'kh A.G., Koonin E.V., Kozayavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).

CC -1- FUNCTION: Is responsible for channeling the electrons from the
CC oxidation of dihydroorotate from the FMN redox center in the pyrd
CC submit to the ultimate electron acceptor NAD(+) (By similarity).
CC -1- COFACTOR: Binds a 2Fe-2S cluster and FAD (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC -1- SUBUNIT: Heterotetramer of 2 pyrk and 2 pyrd subunits (By
CC similarity).

CC -1- SIMILARITY: BELONGS TO THE PYRK FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

CC EMBL; AE010350; AA01779.1; -
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW Pyrimidine biosynthesis; Transport; Electron transport; Metal-binding;
FT Iron; Iron-sulfur; Flavoprotein; FAD; Complete proteome.
FT METAL 217 217 IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
FT METAL 222 222 IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
FT METAL 225 225 IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
FT METAL 234 234 IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 262 AA; 28783 MW; 184A7DCE9B6761 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 262;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 TGVAPLV 53
    |||||
Db 109 TGVAPLV 115

RESULT 18
MTRC_METTH STANDARD; PRT; 266 AA.
ID MTRC_METTH
AC 027229;
DT 15-JUL-1999 (rel. 38, Created)
DT 15-JUL-1999 (rel. 38, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Tetrahydromethanopterin S-methyltransferase subunit C (EC 2.1.1.86)
DE (N5-methyltetrahydromethanopterin--coenzyme M methyltransferase
DE subunit C).
GN MTRC OR MTH1161.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA delta: functional analysis and comparative genomics.";
RT J. Bacteriol. 179:7135-7155(1997).
RL
CC -1- FUNCTION: THIS ENZYME COMPLEX CATALYZES AN INTERMEDIATE STEP IN
CC METHANOGENESIS, THE FORMATION OF METHYL-COENZYME M AND
CC TETRAHYDROMETHANOPTERIN FROM COENZYME M AND N5-METHYL-
CC TETRAHYDROMETHANOPTERIN. THIS IS A ENERGY-CONSERVING, SODIUM-ION
CC TRANSLOCATING STEP (BY SIMILARITY).
CC CATALYTIC ACTIVITY: 5-methyl-5,6,7,8-tetrahydromethanopterin + 2-
CC mercaptoethanesulfonate = 5,6,7,8-tetrahydromethanopterin + 2-
CC (methylthio)ethanesulfonate.
CC PATHWAY: Methanogenesis.
CC -1- SUBUNIT: COMPOSED OF 8 DIFFERENT SUBUNITS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000885; AAB85650.1; -
DR TIGRPMAS; TIGR01148; mtrc; 1.
KW Complete proteome; Transmembrane; Methanogenesis;

```

```

FT INIT_MET 0 0 BY SIMILARITY.
FT TRANSMEM 17 37 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 137 157 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
SQ SEQUENCE 266 AA; 26843 MW; 39B44F59344CDB9F CRC64;

Query Match 4.1%; Score 7; DB 1; Length 266;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALSVL 17
    |||||
Db 140 AALSVL 146

RESULT 19
MTRC_METTH STANDARD; PRT; 266 AA.
ID MTRC_METTH
AC P80185; Q59583;
DT 01-JUL-1993 (rel. 26, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Tetrahydromethanopterin S-methyltransferase subunit C (EC 2.1.1.86)
DE (N5-methyltetrahydromethanopterin--coenzyme M methyltransferase
DE subunit C).
GN MTRC.
OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=79929;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94039026; PubMed=8223548;
RA Stupperich E., Juza A., Hoppert M., Mayer F.;
RT "Cloning, sequencing and immunological characterization of the
RT N5-methyltetrahydromethanopterin: coenzyme-M methyltransferase from
RT Methanobacterium thermoautotrophicum.";
RL Eur. J. Biochem. 217:115-121(1993).
RN [2]
RP SEQUENCE OF 1-30.
RX MEDLINE=93238732; PubMed=8477726;
RA Geertner P., Ecker A., Fischer R., Linder D., Fuchs G., Thauer R.K.;
RT "Purification and properties of N5-
RT methyltetrahydromethanopterin:coenzyme M methyltransferase from
RT Methanobacterium thermoautotrophicum.";
RL Eur. J. Biochem. 213:537-545(1993).
CC -1- FUNCTION: THIS ENZYME COMPLEX CATALYZES AN INTERMEDIATE STEP IN
CC METHANOGENESIS, THE FORMATION OF METHYL-COENZYME M AND
CC TETRAHYDROMETHANOPTERIN FROM COENZYME M AND N5-METHYL-
CC TETRAHYDROMETHANOPTERIN. THIS IS A ENERGY-CONSERVING, SODIUM-ION
CC TRANSLOCATING STEP.
CC CATALYTIC ACTIVITY: 5-methyl-5,6,7,8-tetrahydromethanopterin + 2-
CC mercaptoethanesulfonate = 5,6,7,8-tetrahydromethanopterin + 2-
CC (methylthio)ethanesulfonate.
CC PATHWAY: Methanogenesis.
CC -1- SUBUNIT: COMPOSED OF 8 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X73123; CAA51554.1; -

```

DR PIR: S30343; S30343.
 DR TIGR: TIGR0148; mtic; 1.
 KW Transferase; Methyltransferase; Transmembrane; Methanogenesis.
 FT INIT MET 0
 FT TRANSMEM 17 37 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 75 95 POTENTIAL.
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 137 157 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 208 228 POTENTIAL.
 FT TRANSMEM 229 249 POTENTIAL.
 FT CONFLICT 30 30 Y -> S (IN REF. 2).
 SQ SEQUENCE 266 AA; 26986 MW; 2984360328207807 CRC64;

Query Match
 Best Local Similarity 4.1%; Score 7; DB 1; Length 266;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALSVL 17
 |||||
 DB 140 AAALSVL 146

RESULT 20
 HXDB_HUMAN STANDARD; PRT; 290 AA.
 AC P13378;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-D8 (Hox-4E) (Hox-5.4).
 GN HOXD8 OR HOXA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Birren B., Linton L., Nusbaum C., Lander E.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RA Kosaki K., Kosaki R., Suzuki T., Yoshitashi H., Sasaki K., Matsuo N.;
 RL "A complete mutation analysis panel of human HOX genes";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE OF 193-287 FROM N.A.
 RX MEDLINE=89306602; PubMed=2568311;
 RA Oliver G., Sidell N., Fiske N., Heinzmann C., Mohandas T.,
 RA Sparkes R.S., de Robertis E.M.;
 RL "Complementary homeo protein gradients in developing limb buds";
 RL Genes Dev. 3:641-650(1989).
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC -1- SUBCELLULAR LOCATION: Nucleus.
 CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AC009336; -; NOT ANNOTATED_CDS.
 DR EMBL: AY014304; AAG42152.1;
 DR EMBL: AY014303; AAG42152.1; JOINED.
 DR EMBL: X15507; CAA33529.1;
 DR PIR: B32830; B32830.
 DR PIR: S05957; S05957.

DR HSP: P02833; 9ANT.
 DR TRANSFAC; T03332;
 DR Gene; HGNC:5139; HOXD8.
 DR MIM: 142985;
 DR InterPro: IPR001827; Antennapedia.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR PRINTS: PR00025; ANTENNAPEDIA.
 DR ProDom: PD000010; Homeobox.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00032; HOMEBOX 1; 1.
 DR PROSITE: PS00074; ANTENNAPEDIA; 1.
 DR PROSITE: PS00074; HOMEBOX 2; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 15 23 POLY-ALA.
 FT DOMAIN 45 50 POLY-ALA.
 FT DOMAIN 109 123 POLY-PRO.
 FT DNA_BIND 197 256 HOMEBOX.
 FT CONFLICT 287 287 G -> A (IN REF. 3).
 SQ SEQUENCE 290 AA; 31910 MW; 75FF95A73E2A85F CRC64;

Query Match
 Best Local Similarity 4.1%; Score 7; DB 1; Length 290;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
 |||||
 DB 45 AAAAAL 51

RESULT 21
 PSBO_CHURE STANDARD; PRT; 291 AA.
 AC P12853;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Oxygen-evolving enhancer protein 1, chloroplast precursor (OEE1).
 GN PSBO.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=137C / CC-125;
 RA Mayfield S.P., Schlimer-Rahire G., Frank H., Zuber H., Rochaix J.-D.;
 RL "Analysis of the genes of the OEE1 and OEE3 proteins of the
 RL Photosystem II complex of Chlamydomonas reinhardtii";
 RL Plant Mol. Biol. 12:683-693(1989).
 CC -1- FUNCTION: STABILIZES THE MANGANESE CLUSTER WHICH IS THE PRIMARY
 CC SITE OF WATER SPLITTING (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane; associated
 CC with the photosystem II complex.
 CC -1- SIMILARITY: BELONGS TO THE PSBO FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X13826; CAA32053.1;
 DR PIR: S05508; S05508.
 DR InterPro: IPR002628; PSII_MSP.
 DR Pfam: PF01716; MSP; 1.
 KW Photosynthesis; Photosystem II; Chloroplast; Transit peptide;
 KW Thylakoid; Membrane; Manganese.
 FT TRANSIT 1 52 CHLOROPLAST.

```

FT CHAIN 53 291 OXYGEN-EVOLVING ENHANCER PROTEIN 1.
SQ SEQUENCE 291 AA; 30522 MW; 5DF4DFAB1324267 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
   |||||
Db 34 AAAAAL 40

RESULT 22
RSK_BACHD STANDARD; PRT; 294 AA.
ID RSK_BACHD
AC O9K61;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribokinase (EC 2.7.1.15).
GN RBSK OR BH3728.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 /JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: ATP + D-ribose -> ADP + D-ribose 5-phosphate.
CC -1- PATHWAY: Ribose metabolism; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AP001519; BAB07447.1; -.
CC HSSP: P05054; 1RK2.
CC InterPro: IPR002173; PfKB.
CC InterPro: IPR002139; Ribokinase.
CC Pfam: PF00294; PFKB; 1.
CC PRINTS: PR00990; RIBOKINASE.
CC PROSITE: PS00583; PFKB_KINASES_1; 1.
CC PROSITE: PS00584; PFKB_KINASES_2; 1.
CC Transferase; Kinase; Complete proteome.
SQ SEQUENCE 294 AA; 31089 MW; 8C13E0FCF5E89FDE CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAALSV 16
   |||||
Db 264 AAAALSV 270

RESULT 23
GSH2_HUMAN STANDARD; PRT; 304 AA.
ID GSH2_HUMAN
AC G9BZM3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein GSH-2.
GN GSH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Sakai T., Sakamoto S., Nakamura K., Muraki T.;
RT "Human homeobox protein GSH-2.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cools J., Marynen P.;
RT "The sequence of the human GSH2 gene.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA
CC SEQUENCE 5'-CNAATTAG-3' (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AB028638; BAB84822.1; -.
CC EMBL: AF306344; AAK00880.1; -.
CC EMBL: AF306343; AAK00880.1; JOINED.
CC HSSP: P14653; 1B72.
CC InterPro: IPR000047; HTH_repressor.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 1.
CC PRINTS: PR00024; HOMEOBOX.
CC PRINTS: PR00031; HTHREPRESSR.
CC PRODOM: PD000010; Homeobox; 1.
CC SMART: SM00389; HOX; 1.
CC PROSITE: PS00027; HOMEOBOX_1; 1.
CC PROSITE: PS00027; HOMEOBOX_2; 1.
CC Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
CC Developmental protein.
CC DNA_BIND 202 261 HOMEOBOX.
FT DOMAIN 124 130 POLY-HIS.
FT DOMAIN 134 139 POLY-HIS.
FT DOMAIN 147 162 POLY-ALA.
SQ SEQUENCE 304 AA; 32061 MW; E896D54224886C1 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
   |||||
Db 157 AAAAAL 163

RESULT 24
GSH2_MOUSE STANDARD; PRT; 305 AA.
ID GSH2_MOUSE
AC P31316;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein GSH-2.
GN GSH2 OR GSH-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIH SWISS;
RX MEDLINE=9534493; PubMed=7619729;
RA Hsieh-Li H.M., Witte D.P., Stancik J.C., Weinstein M., Li H.,
RT Porter S.S.;
RL "Gsh-2, a murine homeobox gene expressed in the developing brain.";
RM Mech. Dev. 50:177-186(1995).
[2]
RP SEQUENCE OF 203-262 FROM N.A.
RX MEDLINE=9207356; PubMed=1683707;
RA Singh G., Kaur S., Stock J.L., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Potter S.S.;
RL "Identification of 10 murine homeobox genes.";
RM Proc. Natl. Acad. Sci. U.S.A. 88:10706-10710(1991).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR THAT BINDS TO THE DNA
CC SEQUENCE 5'-CNAATTAG-3'.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING BRAIN.
CC -1- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S79041; AAB34947.1; -
DR PIR: B37290; B37290.
DR PIR: B38809; B38809.
DR HSSP: P14653; 1B72.
DR MGD: MGT:95843; Gsh2.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PR00024; HOMEOBOX.
DR ProDom: PD00010; HOMEOBOX.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PS00027; HOMEOBOX_1; 1.
DR PROSITE: PS50071; HOMEOBOX_2; 1.
KW Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
KW Developmental protein.
FT DNA_BIND 203 262 HOMEOBOX.
FT DOMAIN 124 130 POLY-HIS.
FT DOMAIN 134 139 POLY-HIS.
FT DOMAIN 147 163 POLY-ALA.
SQ SEQUENCE 305 AA; 32167 MW; 51E7F2DB76E32608 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 305;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 158 AAAAAA 164

```

```

RC STRAIN-DSM 1411;
RX MEDLINE=93021102; PubMed=1404376;
RA Engert C., Krueger K., Offner S., Pfeifer F.;
RL "Three different but related gene clusters encoding gas vesicles in
halophilic archaea.";
RM J. Mol. Biol. 227:586-592(1992).
CC -1- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS
CC VESICLE SYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64701; CAA5954.1; -
DR PIR: S28125; S28125.
DR Gas vesicle.
SQ SEQUENCE 322 AA; 36382 MW; EEDB60511E10EE5 CRC64;

Query Match 4.1%; Score 7; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 164 QNRRAEL 170
DB 234 QNRRAEL 240

```

```

RESULT 26
ID SLPI_DROME STANDARD; PRT; 322 AA.
AC P32030; Q9YQV4;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fork head domain transcription factor slp1 (Slippy paired locus
protein 1).
GN SLPI OR FDB OR CG16738.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Canton-S, and Oregon-R;
RX MEDLINE=92275347; PubMed=1317319;
RA Grossniklaus U., Pearson R.K., Gehring W.J.;
RL "The Drosophila sloppy paired locus encodes two proteins involved in
segmentation that show homology to mammalian transcription factors.";
RM Genes Dev. 6:1030-1051(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Chang M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck B., Brooksstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,

```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., Moshneff A.,
 RA Mouton S.M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weltsch G.M., Welschbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN SEGMENTATION. REQUIRED
 CC FOR THE FORMATION OF THE MANDIBULAR LOBE. DIFFERENT LEVELS OF SLIP
 CC ACTIVITY SEEM TO BE REQUIRED IN DIFFERENT SEGMENTS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE POSTERIOR HALF OF EACH
 CC PARASEGMENT JUST ANTERIOR TO THE PARASEGMENTAL BOUNDARY.
 CC -1- DEVELOPMENTAL STAGE: PRESENT AT 0-3 HRS OF EMBRYOGENESIS. MAXIMAL
 CC EXPRESSION AT 3-6 HRS. STRONG RE-EXPRESSION IN FIRST-INSTAR
 CC LARVAE.
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X66095; CAA46889.1; -;
 DR EMBL: X66096; CAA46890.1; -;
 DR EMBL: AE003578; AAF51058.1; -;
 DR PIR: S23053; S23053.
 DR PIR: S23054; S23054.
 DR HSSP: Q63245; 2HRF.
 DR TRANSFAC: T01055; -;
 DR FLYbase: FBgn0003430; slp1.
 DR InterPro: IPR001766; TF_Fork_head.
 DR Pfam: PF00250; Fork_head; 1.
 DR PRINTS: PR00053; FORKHEAD.
 DR PRODOM: PD000425; TF_Fork_head; 1.
 DR SMART: SM00339; FH; 1.
 DR PROSITE: PS00657; FORK_HEAD_1; 1.
 DR PROSITE: PS00658; FORK_HEAD_2; 1.
 DR PROSITE: PS00659; FORK_HEAD_3; 1.
 KW DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation; Pair-rule protein; Polymorphism.
 FT DNA_BIND 119 210 FORK-HEAD.
 FT DOMAIN 256 262 POLY-ALA.
 FT VARIANT 246 246 P -> S (IN STRAIN OREGON-R AND BERKELEY).
 FT VARIANT 303 303 P -> Q (IN STRAIN OREGON-R AND BERKELEY).
 FT SEQUENCE 322 AA; 36202 MW; 24CDE9F0102024C4 CRC64;
 SQ
 Query Match 4.1%; Score 7; DB 1; Length 322;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 27
 ID A85B_MYCKA STANDARD; PRT; 325 AA.
 AC A85B_MYCKA
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antigen 85-B precursor (85B) (Extracellular alpha-antigen) (Antigen 85
 DE complex B) (A85B) (Mycolyt transferase 85B) (EC 2.3.1.-)
 DE (Fibronectin-binding protein B).
 GN FBPB.
 OS Mycobacterium kansasii.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1768;
 RX MEDLINE=90129315; PubMed=2404875;
 RP SEQUENCE FROM N.A.
 RA Matsuo K., Yamaguchi R., Yamazaki A., Tasaka H., Terasaka K.,
 RA Yamada T.;
 RT "Cloning and expression of the gene for the cross-reactive alpha
 RT antigen of *Mycobacterium kansasii*.";
 RL Infect. Immun. 58:550-556(1990).
 CC -1- FUNCTION: PROTEINS OF THE ANTIGEN 85 COMPLEX ARE RESPONSIBLE FOR
 CC THE HIGH AFFINITY OF MYCOBACTERIA TO FIBRONECTIN. POSSESSES A
 CC MYCOLYTIC TRANSFERASE ACTIVITY REQUIRED FOR THE BIOSYNTHESIS OF
 CC TREHALOSE DIMYCOLATE (CORD FACTOR), A DOMINANT STRUCTURE NECESSARY
 CC FOR MAINTAINING CELL WALL INTEGRITY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: TO OTHER MYCOBACTERIUM A85 ANTIGENS AND TO THE
 CC N-TERMINAL OF C. GLUTAMICUM P51 PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X53897; CAA37868.1; -;
 DR PIR: A37185; A37185.
 DR HSSP: P31953; 1DOY.
 DR InterPro: IPR000801; Esterase_put.
 DR Pfam: PF00756; Esterase; 1.
 KW Transferase; Acyltransferase; Antigen; Signal.
 FT SIGNAL 1 40
 FT CHAIN 41 325 ANTIGEN 85-B.
 FT ACT_SITE 166 166 BY SIMILARITY.
 FT ACT_SITE 270 270 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 FT SEQUENCE 325 AA; 34323 MW; 5F2281BCC48AE30D CRC64;
 SQ
 Query Match 4.1%; Score 7; DB 1; Length 325;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 DB 257 AAAAAA 263

GN STC25A16 OR GDA OR GDC.
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=93091248; PubMed=1457817;
 RA Fiermonte G., Runswick M.J., Walker J.E., Palmieri F.;
 RT "Sequence and pattern of expression of a bovine homologue of a human
 mitochondrial transport protein associated with Grave's disease.";
 RL DNA Seq. 3:71-78(1992).
 CC
 CC -1- FUNCTION: Required for the accumulation of coenzyme A in the
 CC mitochondrial matrix (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane (By similarity).
 CC -1- TISSUE SPECIFICITY: MOSTLY IN THYROID, LIVER, LUNG, KIDNEY AND
 CC TO A LESSER EXTENT IN HEART AND SKELETAL MUSCLE.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X66035; CAA46834.1; -
 DR PIR: S26596; S26596;
 DR InterPro: IPR002067; Mit-carrier.
 DR InterPro: IPR001993; Mitoch-carrier.
 DR Pfam: PF00153; mito_carrier; 3.
 DR PRINTS: PR00926; MITOCH_CARRIER; 3.
 DR PROSITE: PS00215; MITOCH_CARRIER; 3.
 KW Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 189 209 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT REPEAT 1 121 1.
 FT REPEAT 122 217 2.
 FT REPEAT 218 330 3.
 SQ SEQUENCE 330 AA; 36085 MW; 4C614701D2B8DEA9 CRC64;
 Query Match
 Best Local Similarity 4.1%; Score 7; DB 1; Length 330;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 AAAAAAAL 14
 Db 2 AAAAAAL 8
 RESULT 29
 ID RLAO_METUA STANDARD; PRT; 338 AA.
 AC PS4049;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Acidic ribosomal protein P0 homolog (L10E).
 GN RPLP0 OR M05059.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 CC NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utermack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL Science 273:1058-1073(1996).
 CC
 CC -1- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
 CC OF E. COLI PROTEIN L10.
 CC
 CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U67500; AAB8499.1; -
 DR TIGR: M05059;
 DR InterPro: IPR001790; Ribosomal_L10.
 DR Pfam: PF00466; Ribosomal_L10; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 338 AA; 36751 MW; 63A6ARD357E3052D CRC64;
 Query Match
 Best Local Similarity 4.1%; Score 7; DB 1; Length 338;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 61 ALASKLP 67
 Db 282 ALASKLP 288
 RESULT 30
 ID GUB_FIBSU STANDARD; PRT; 349 AA.
 AC P17989;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Beta-glucanase precursor (EC 3.2.1.73) (Endo-beta-1,3-1,4 glucanase)
 DE (1,3-1,4-beta-D-glucan 4-glucanohydrolase) (Mixed linkage beta-
 DE glucanase) (Lichenase).
 OS Fibrobacter succinogenes (Bacteroides succinogenes).
 CC Bacteria; Fibrobacter/Acidobacteria group; Fibrobacter group;
 CC Fibrobacter.
 CC NCBI_TaxID=833;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 28-57.
 RC STRAIN=Isolate S85;
 RX MEDLINE=90299807; PubMed=2193918;
 RA Teather R.M., Efrile J.D.;
 RT "DNA sequence of a Fibrobacter succinogenes mixed-linkage
 beta-glucanase (1,3-1,4-beta-D-glucan 4-glucanohydrolase) gene";
 RL J. Bacteriol. 172:3837-3841(1990).
 CC
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in beta-D-glucans containing 1,3- and 1,4-bonds.
 CC
 CC -1- SIMILARITY: BELONGS TO FAMILY 16 OF GYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M33676; AAA24896.1; -
 DR PIR; A44507; A44507.
 DR HSSP; P23904; 1AKK.
 DR InterPro; IPR000757; glyco_hydro_16.
 DR Pfam; PF00722; glyco_hydro_16; 1.
 DR PROSITE; PS01034; GLYCOSYL_HYDROL_F16; 1.
 KW Hydrolyase; Glycosidase; Signal; Repeat.
 FT SIGNAL 1 27
 FT CHAIN 28 349
 FT ACT_SITE 79 79 BETA-GLUCANASE.
 FT ACT_SITE 83 83 NUCLEOPHILE (BY SIMILARITY).
 FT DOMAIN 271 307 PROTON DONOR (BY SIMILARITY).
 FT REPEAT 271 277 5 x 7 AA TANDEM REPEATS OF P-X-S-S-S-X.
 FT REPEAT 271 277 1.
 FT REPEAT 284 284 2.
 FT REPEAT 285 291 3.
 FT REPEAT 292 298 4.
 FT REPEAT 301 307 5.
 FT REPEAT 307 307 16DC4F5BDEFC578A CRC64;
 FT SEQUENCE 349 AA; 37737 MW; 16DC4F5BDEFC578A CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 349;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 AAAAAL 14
 Db 15 AAAAAL 21
 RESULT 31
 DKK3_HUMAN STANDARD; PRT; 350 AA.
 AC Q9UBP4; Q9UBP4.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dickkopf related protein-3 precursor (Dkk-3) (Dkkopf-3) (Dkk-3).
 GN DKK3 OR REIC.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Petal brain.
 RC MEDLINE=20035735; PubMed=10570958;
 RA Kiripunk V.E., Sharp J.D., Jiang C., Robison K., Chickering T.W.,
 RA Amaravadi L., Brown D.E., Guyot D., Mays G., Leiby K., Chang B.,
 RA Duong T., Goodearl A.D.J., Gearing D.P., Sokol S.Y., McCarthy S.A.;
 RT "Functional and structural diversity of the human Dickkopf gene
 RT family";
 RT Gene 238:301-313(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tanaka S., Sugimachi K., Sugimachi K.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20119095; PubMed=10652205;
 RA Tsuji T., Miyazaki M., Sakaguchi M., Inoue Y., Namba M.;
 RT "A REIC gene shows down-regulation in human immortalized cells and
 RT human tumor-derived cell lines";
 RL Biochem. Biophys. Res. Commun. 268:20-24(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Tate G., Mitsuya T.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC PubMed=11814687;
 RA Kobayashi K., Ouchida M., Tsuji T., Hanafusa H., Miyazaki M.,
 RA Namba M., Shimizu N., Shimizu K.;
 RT "Reduced expression of the REIC/Dkk-3 gene by promoter-
 RT hypermethylation in human tumor cells";

RL Gene 282:151-158(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Kidney;
 RA Straube R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INHIBITOR OF WNT SIGNALING PATHWAY (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN HEART, BRAIN, AND SPINAL
 CC CORD.
 CC -1- PTM: N-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO THE DICKKOPF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF177396; AAF02676.1; -
 DR EMBL; AB033421; BAA85488.1; -
 DR EMBL; AB034203; BAA80548.1; -
 DR EMBL; AB035182; BAA87044.2; -
 DR EMBL; AB045205; BAA87044.2; JOINED.
 DR EMBL; AB045206; BAA87044.2; JOINED.
 DR EMBL; AB045207; BAA87044.2; JOINED.
 DR EMBL; AB045208; BAA87044.2; JOINED.
 DR EMBL; AB045209; BAA87044.2; JOINED.
 DR EMBL; AB045210; BAA87044.2; JOINED.
 DR EMBL; AB057591; BAA84360.1; -
 DR EMBL; AB057804; BAA84361.1; -
 DR EMBL; BC007660; AAB07660.1; -
 DR Gene; HGNC:2893; DKK3.
 DR MIM; 605416; -
 KW Developmental protein; Signal; Glycoprotein.
 FT SIGNAL 1 16
 FT CHAIN 17 350
 FT DOMAIN 147 195 DICKKOPF RELATED PROTEIN-3.
 FT DOMAIN 208 284 DKK-TYPE CYS-2.
 FT DOMAIN 338 343 POLY-ALA.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 335 335 G -> R (IN REF. 4).
 FT SEQUENCE 350 AA; 38291 MW; 72F504122B40AFEE CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 350;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 AAAAAL 14
 Db 338 AAAAAL 344
 RESULT 32
 KLF2_MOUSE STANDARD; PRT; 354 AA.
 ID KLF2_MOUSE
 AC O60843;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Kruppel-like factor 2 (Lung kruppel-like factor).
 DE Kruppel-like factor 2 (Lung kruppel-like factor).
 GN KLF2 OR KLF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

CC STRAIN-C57BL/6 X CBA; TISSUE=Lung;
 RA MEDLINE=96025976; PubMed=7565748;
 RT Anderson K.P., Kern C.B., Crabbe S.C., Lingrel J.B.;
 RT "Isolation of a gene encoding a functional zinc finger protein
 RT homologous to erythroid Kruppel-like factor: identification of a new
 RT multigene family.";
 RL Mol. Cell. Biol. 15:5957-5965(1995).
 CC -1- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
 CC AND ACTIVATES TRANSCRIPTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN THE LUNGS AND
 CC SPLEEN.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U25096; AAA86728.1; -
 DR HSSP: P08047; 1SP2.
 DR TRANSFAC: T01677; -
 DR MGD: MGI:134272; Klf2.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00006; Znf_C2H2_3.
 DR PRINTS: PR00048; ZNCFINGER.
 DR PRODOM: PD000003; Znf_C2H2; 2.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
 DR Transcription regulation: Activator; Zinc-finger; Metal-binding;
 KW DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 60 70 POLY-PRO.
 FT 165 169 POLY-PRO.
 FT 224 230 POLY-ALA.
 FT 271 353 ZINC_FINGERS.
 FT 271 295 C2H2-TYPE.
 FT 301 325 C2H2-TYPE.
 FT 331 353 C2H2-TYPE.
 SQ SEQUENCE 354 AA; 37700 MW; CAA99D018AC5BAF7 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 354;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 AAAAAL 14
 DB 225 AAAAAL 231
 RESULT 33
 KLF2_HUMAN STANDARD; PRT; 355 AA.
 AC Q9Y5W3; Q9UKR6; Q9U5S5;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Kruppel-like factor 2 (Lung Kruppel-like factor).
 GN KLF2 OR KLF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99231781; PubMed=10217429;
 RA Kozyrev S.V., Hansen L.L., Poltarau A.B., Domlinsky D.A.,
 RA Kisselev L.L.;

RT "Structure of the human Cpg-island-containing lung Kruppel-like factor
 RT (KLF2) gene and its location in chromosome 19p13.11-13 locus.";
 RT FEBS Lett. 448:149-152(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99389728; PubMed=10458913;
 RA Wanl M.A., Conkright M.D., Jeffries S., Hughes M.J., Lingrel J.B.;
 RT "cDNA isolation, genomic structure, regulation, and chromosomal
 RT localization of human lung Kruppel-like factor.";
 RL Genomics 60:78-86(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lee H.J., Kim M.K., Kim Y.H., Seo J.M., Lee H.M., Chung H.J.,
 RA Sohn M.Y., Hwang S.Y., Im S.O., Jung E.J., Kim J.C.;
 RT "A catalogue of genes in the human dermal papilla cells as identified
 RT by expressed sequence tags.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER
 CC AND ACTIVATES TRANSCRIPTION (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF123344; AAD25076.1; -
 DR EMBL: AF134053; AAD55891.1; -
 DR EMBL: AF205849; AAF13295.1; -
 DR HSSP: P08047; 1SP2.
 DR TRANSFAC: T04958; -
 DR Genew: HGNC:6347; KLF2.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00006; Znf_C2H2; 3.
 DR PRODOM: PD000003; Znf_C2H2; 2.
 DR SMART: SM00355; Znf_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
 KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
 KW DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 62 71 POLY-PRO.
 FT 130 135 POLY-GLY.
 FT 167 171 POLY-PRO.
 FT 225 231 POLY-ALA.
 FT 272 354 ZINC_FINGERS.
 FT 272 296 C2H2-TYPE.
 FT 302 326 C2H2-TYPE.
 FT 332 354 C2H2-TYPE.
 FT 43 43 S -> N (IN REF. 2).
 FT 104 104 L -> P (IN REF. 1).
 FT 175 175 P -> S (IN REF. 2).
 FT 184 184 L -> M (IN REF. 2).
 SQ SEQUENCE 355 AA; 37419 MW; D5849C831D676AE1 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 355;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 AAAAAL 14
 DB 226 AAAAAL 232
 RESULT 34
 BET3_MESAU STANDARD; PRT; 367 AA.
 ID BET3_MESAU

AC 009029; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE BERR3 protein.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96140430; PubMed=8552091;
 RA Peyton M., Stellrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,
 RA Tsai M.-J.;
 RA "BERR3, a novel helix-loop-helix protein, can act as a negative
 RT regulator of BERR2 and MyoD-responsive genes.";
 RT Mol. Cell. Biol. 16:626-633(1996).
 RL [1]
 CC -1- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3
 CC (E47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF
 CC NEUROD1 AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION
 CC WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE
 CC OF AN INTRACT BASIC DOMAIN, DOES NOT BIND TO DNA.
 CC -1- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA).
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "ATOMAL" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: S80870; AAB50691.1; -;
 DR TRASNFRAC: T01674; -;
 DR InterPro: IPR001092; HLH-basic.
 DR Pfam: PF00010; HLH: 1.
 DR SMART: SM00353; HLH: 1.
 DR PROSITE: PS00038; HLH_2; 1.
 DR PROSITE: PS50888; HLH_2; 1.
 KW Nuclear protein; transcription regulation; Repressor.
 FT DOMAIN 11 14
 FT DOMAIN 58 62 POLY-ALA.
 FT DOMAIN 62 99 POLY-SER.
 FT DOMAIN 174 179 POLY-GLY.
 FT DOMAIN 204 217 POLY-GLY.
 FT DNA_BIND 229 240 POLY-GLY.
 FT DNA_BIND 241 282 BASIC DOMAIN.
 FT DOMAIN 311 319 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT POLY-ALA.
 SQ SEQUENCE 367 AA: 35905 MW: 60AB9AF96E85F77 CRC64;
 Query Match 4.1%; Score 7; DB 1; Length 367;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Kumano M., Tamakoshi A., Yamane K.;
 RT "A 32 kb nucleotide sequence from the region of the *lincomycin*-
 RT resistance gene (22-25 degree) of the *Bacillus subtilis* chromosome and
 RT identification of the site of the *lin-2* mutation."
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berteiro M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignelli S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Goffeau A., Goldbly E.J., Grandi G.,
 RA Ghim S.Y., Glaser P., Goffeau A., Hachez K., Hachez D.,
 RA Guisepi G., Guy B.J., Haga K., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Hilbert H., Holsappel S., Hosono S., Kiehl-Blanchard M., Klein C.,
 RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Portetelle S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solio B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Tarpstra P., Tognoni A.,
 RA Tosiato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzmeyer T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 RL [3]
 RP SEQUENCE OF 1-319 FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=95219079; PubMed=7704254;
 RA Ogasawa K., Akagawa E., Nakamura K., Yamane K.;
 RT "Determination of a 21548 bp nucleotide sequence around the 24
 RT degrees region of the *Bacillus subtilis* chromosome."
 RL Microbiology 141:269-275(1995).
 RN [4]
 RP SEQUENCE OF 1-147 FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=92339527; PubMed=1353026;
 RA Awade A., Cleuzat P., Gonzales T., Robert-Baudouy J.;
 RT "Characterization of the *pcp* gene encoding the pyrrolidone carboxyl
 RT peptidase of *Bacillus subtilis*.";
 RL FEBS Lett. 305:67-73(1992).
 CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to two
 CC frameshifts.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch.

CC EMBL: AB000617; BAA22227.1; -
 CC EMBL: 299105; CAB12060.1; -
 DR EMBL: D30808; -; NOT_ANNOTATED_CDS.
 DR EMBL: X66034; -; NOT_ANNOTATED_CDS.
 DR Subtilisin; Bg11176; ycbu.
 DR InterPro: IPR00192; AminoTransfV.
 DR Pfam: PF00266; aminotran_5; 1.
 DR PROSITE: PS00595; AA_TRANSFER_CLASS_5; 1.
 KW Hypothetical protein; Signal; Transferase; AminoTransferase;
 KW Pyridoxal phosphate; Complete proteome.
 FT SIGNAL 1 30
 FT CHAIN 1 30
 FT BINDING 207 207
 FT CONFLICT 92 92
 FT CONFLICT 133 133
 FT CONFLICT 319 319
 SQ SEQUENCE 370 AA; 40809 MM; 38C91AF0FC27DF53 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 370;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 AAAAALS 15
 Db 265 AAAAALS 271

RESULT 36
 FL_ORYSA STANDARD; PRT; 389 AA.
 AC 024175;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative transcription factor FL (RFL).
 GN FL.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC NCBT_TaxID=4530;
 RA Kiyozuka J., Konishi S., Nemoto K., Izawa T., Shimamoto K.;
 RT "Down-regulation of RFL, the FLO/LFY homolog of rice, accompanied
 with panicle branch initiation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:1979-1982(1998).
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: In very young panicle but not in mature
 florets, mature leaves, roots or apical meristems.
 CC -1- SIMILARITY: BELONGS TO THE FLO / LFY FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AB005620; BAA21547.1; -
 DR InterPro: IPR002910; FLO_LFY; 1.
 DR Pfam: PF01698; FLO_LFY; 1.
 KW Transcription regulation; Activator; DNA-binding;
 KW Nuclear protein; Developmental protein.
 FT DOMAIN 22 37
 FT DOMAIN 201 204
 FT DOMAIN 213 217

POLY-PRO.
 POLY-GLY.
 POLY-GLY.

FT DOMAIN 378 383
 SQ SEQUENCE 389 AA; 42530 MM; 5132AA6A34B0868A CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 389;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
 Db 378 AAAAAL 384

RESULT 37
 HB9_HUMAN STANDARD; PRT; 401 AA.
 AC P50219;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein HB9.
 GN HLXB9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBT_TaxID=9606;
 RA Harrison K.A., Druey K.M., Deguchi Y., Tuscano J.M., Kehrl J.H.;
 RT "A novel human homeobox gene distantly related to proboscipedia is
 expressed in lymphoid and pancreatic tissues.";
 RL J. Biol. Chem. 269:19968-19975(1994).
 CC -1- FUNCTION: PUTATIVE TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND PANCREATIC TISSUES.
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U07664; AAB60647.1; -
 DR EMBL: U07663; AAB60647.1; JOINED.
 DR HSSP: P14653; 1872.
 DR TRANSFAC: T03420; -
 DR GeneW: HGNC:4979; HLXB9.
 DR MIM: 142994; -
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR PRINTS: PR00024; HOMEBOX.
 DR PRODOM: PD000010; Homeobox; 1.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Nuclear protein; Transcription regulation.

FT DOMAIN 39 48
 FT DOMAIN 97 111
 FT DOMAIN 120 135
 FT DOMAIN 169 177
 FT DNA_BIND 242 301
 FT DOMAIN 316 325
 SQ SEQUENCE 401 AA; 40932 MM; 0006AED71D594FE CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 1; Length 401;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14

Db 170 AAAAAL 176

RESULT 38
NU4M_ASCSU STANDARD: PRT: 409 AA.
ID NU4M_ASCSU
AC P24880:
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
GN ND4.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC Mitochondrion.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Body wall muscle, and Egg;
RA MEDLINE=92201635; PubMed=1551572;
RX Okamoto R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;
RT "The mitochondrial genomes of two nematodes, Caenorhabditis elegans
and Ascaris suum.";
RL Genetics 130:471-498(1992).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54253; CAA38170.1; -
DR PIR: S26021; S26021.
DR InterPro: IPR001750; Oxidored_g1.
DR Pfam: PF00361; Oxidored_g1.1
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KM Oxidoreductase; NAD; Ubiquinone; MA135C1CA558786 CRC64;
SQ SEQUENCE 409 AA; 47064 MW; AA135C1CA558786 CRC64;
SO
Query Match 4.1%; Score 7; DB 1; Length 409;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 68 SLVYDF 74
DB 147 SLVYDF 153
RESULT 39
RL4_HUMAN STANDARD: PRT: 427 AA.
ID RL4_HUMAN
AC P36578; P39029; Q96929;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L4 (L1).
GN RPL4 OR RPL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=94092742; PubMed=8268230;
RX Bagni C., Mariotti P., Annesi F., Amaldi F.;
RT "Human ribosomal protein L4: cloning and sequencing of the cDNA and
primary structure of the protein.";
RL Biochim. Biophys. Acta 1216:475-478(1993).
SO

RN [2]
RP REVISIONS.
RA Bagni C.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoma;
RA Kato S.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC MEDLINE=21864036; PubMed=11875025;
RX Yoshinaka M., Uechi T., Asakawa S., Shimizu N., Kenmochi N.,
RA Maeda N., Minoshima S., Tanaka T., Shimizu N., Kenmochi N.;
RT "The human ribosomal protein genes: sequencing and comparative
analysis of 73 genes.";
RL Genome Res. 12:379-390(2002).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, Eye, Muscle, Pancreas, and Skin;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP -1- SIMILARITY: BELONGS TO THE LAE FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73974; CAA52154.1; -
DR EMBL: L20868; AAA60281.1; ALT-SEQ.
DR EMBL: D23660; BAA04887.1; -
DR EMBL: BC001365; AAH01365.1; -
DR EMBL: BC005817; AAH05817.1; -
DR EMBL: BC007259; AAH07259.1; -
DR EMBL: BC007748; AAH07748.1; -
DR EMBL: BC007996; AAH07996.1; -
DR EMBL: BC009888; AAH09888.1; -
DR EMBL: BC010151; AAH10151.1; -
DR EMBL: BC014653; AAH14653.1; -
DR EMBL: AB061820; BAB79458.1; -
DR PIR: S37197; S37197.
DR PIR: S39803; S39803.
DR SWISS-2DPAGE: P36578; HUMAN.
DR GeneW: HGNC:10353; RPL4.
DR MIM: 180479; -
DR InterPro: IPR002136; Ribosomal_L4/L1E.
DR Pfam: PF00573; Ribosomal_L4.1.
DR PROSITE: PS00939; RIBOSOMAL_L1E; 1.
KM Ribosomal protein.
RN [7]
RP DOMAIN 364 427
FT CONFICT 3 3 LYS-RICH.
FT CONFICT 36 36 C->V (IN REF. 1).
FT CONFICT 63 63 I->M (IN REF. 1).
FT CONFICT 147 147 S->R (IN REF. 1).
FT CONFICT 147 147 V->F (IN REF. 1).
FT CONFICT 201 201 MISSING (IN REF. 1).
SQ SEQUENCE 427 AA; 47697 MW; 4785ED31699CD792 CRC64;
SO
Query Match 4.1%; Score 7; DB 1; Length 427;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AAAAAL 14
DB 355 AAAAAL 361
RESULT 40
TUD4_AGRVI STANDARD: PRT: 438 AA.
ID TUD4_AGRVI
SO

AC 044472;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative hydroxypyruvate reductase (EC 1.1.1.81).
 GN TTUD.
 OS Agrobacterium vitis (Rhizobium vitis).
 OG Plasmid pTRAB4.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=373;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB4;
 RX MEDLINE=96062236; PubMed=7592423;
 RA Crouzet P., Otten L.;
 RT "Sequence and mutational analysis of a tartrate utilization operon
 from Agrobacterium vitis.";
 RL J. Bacteriol. 177:6518-6526(1995).
 CC -1- FUNCTION: DEGRADATES AN UNIDENTIFIED TOXIC PRODUCT FROM THE FIRST
 CC -1- STEP OF TARTRATE DEGRADATION.
 CC -1- CATALYTIC ACTIVITY: D-glycerate + NAD(P)(+) = hydroxypyruvate +
 CC NAD(P)H.
 CC -1- PATHWAY: Tartrate degradation; second step.
 CC -1- INDUCTION: By TARTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U25634; AAA68699.1;
 DR InterPro: IPR005346; UPE0125.
 DR Pfam: PF03658; UPE0125; 1
 DR Plasmid: Oxidoreductase: NADP.
 KW SEQUENCE 438 AA; 44730 MW; 38C75BAFFB0636B CRC64;
 SQ

Query Match
 Best Local Similarity 4.18; Score 7; DB 1; Length 438;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
 |||||
 Db 276 AAAAAL 282

Search completed: July 6, 2003, 14:20:46
 Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:19:30 ; Search time 32 Seconds
(without alignments)
1107.503 Million cell updates/sec

Title: US-09-674-779B-2

Perfect score: 172

Sequence: 1 MMLHIQIAAAALSVLTFM.....IATGTNEANSQNRRAELSY 172

Scoring table: OLIGO

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database:

SPTRMBL_21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mmc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriaphage:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	5.8	186	Q9PC85	Q9PC85 xylella fas
2	9	5.2	998	Q94EA6	Q94EA6 oryza sativ
3	9	5.2	2168	Q9VOM0	Q9VOM0 drosophila
4	9	5.2	2176	Q46112	Q46112 drosophila
5	8	4.7	143	Q9X0G9	Q9X0G9 thermocoga
6	8	4.7	249	Q8S1R2	Q8S1R2 oryza sativ
7	8	4.7	258	Q9B1U7	Q9B1U7 drosophila
8	8	4.7	268	Q93NP5	Q93NP5 escherichia
9	8	4.7	280	Q930C4	Q930C4 rhizobium m
10	8	4.7	306	Q9VVI8	Q9VVI8 rhizobium m
11	8	4.7	316	Q986T0	Q986T0 rhizobium 1
12	8	4.7	324	Q9NGI9	Q9NGI9 drosophila
13	8	4.7	324	Q9NGI8	Q9NGI8 drosophila
14	8	4.7	324	Q9NGI7	Q9NGI7 drosophila
15	8	4.7	324	Q9NGB4	Q9NGB4 drosophila
16	8	4.7	324	Q9N6K3	Q9N6K3 drosophila

17	8	4.7	324	Q9N6K2	Q9N6K2 drosophila
18	8	4.7	381	Q8X1X2	Q8X1X2 aspergillus
19	8	4.7	452	Q9J5Z1	Q9J5Z1 neisseria m
20	8	4.7	504	Q981K5	Q981K5 rhizobium 1
21	8	4.7	526	Q9HTN7	Q9HTN7 pseudomonas
22	8	4.7	541	Q9S7V5	Q9S7V5 arabidopsis
23	8	4.7	545	Q9S511	Q9S511 drosophila
24	8	4.7	547	Q9XAX4	Q9XAX4 pseudomonas
25	8	4.7	570	Q9M7S3	Q9M7S3 lolium pere
26	8	4.7	587	Q9VOP7	Q9VOP7 burkholderi
27	8	4.7	587	Q02427	Q02427 ciona intes
28	8	4.7	624	Q9W2T0	Q9W2T0 drosophila
29	8	4.7	637	Q9W0E6	Q9W0E6 drosophila
30	8	4.7	637	Q8T6B9	Q8T6B9 drosophila
31	8	4.7	644	Q9PRV3	Q9PRV3 oryza sativ
32	8	4.7	644	Q9PRV2	Q9PRV2 oryza sativ
33	8	4.7	651	Q9HKB4	Q9HKB4 thermoplasm
34	8	4.7	662	Q9V9F7	Q9V9F7 drosophila
35	8	4.7	783	Q9QY56	Q9QY56 mus musculu
36	8	4.7	824	Q960R1	Q960R1 drosophila
37	8	4.7	825	Q9BMZ8	Q9BMZ8 drosophila
38	8	4.7	975	Q962D2	Q962D2 drosophila
39	8	4.7	1015	Q98ZW9	Q98ZW9 drosophila
40	8	4.7	1164	Q9Y161	Q9Y161 drosophila
41	8	4.7	1321	Q96ZD1	Q96ZD1 drosophila
42	8	4.7	1323	Q9NMX4	Q9NMX4 drosophila
43	8	4.7	1376	Q9BMZ9	Q9BMZ9 drosophila
44	8	4.7	1412	Q9VKJ1	Q9VKJ1 drosophila
45	8	4.7	1976	Q9WXI8	Q9WXI8 bacillus sp
46	8	4.7	2090	Q9WZT1	Q9WZT1 drosophila
47	8	4.7	2703	Q9VEG7	Q9VEG7 drosophila
48	8	4.7	2715	Q61603	Q61603 drosophila
49	8	4.1	38	Q9S9D9	Q9S9D9 nicotiana t
50	7	4.1	71	Q91LL1	Q91LL1 white spot
51	7	4.1	78	Q9BMP6	Q9BMP6 culicoides
52	7	4.1	82	Q9YH51	Q9YH51 pseudopleur
53	7	4.1	82	Q99013	Q99013 pseudopleur
54	7	4.1	96	Q8W327	Q8W327 oryza sativ
55	7	4.1	105	Q943N0	Q943N0 oryza sativ
56	7	4.1	110	Q942Y1	Q942Y1 oryza sativ
57	7	4.1	114	Q94E39	Q94E39 oryza sativ
58	7	4.1	122	Q9Q1V4	Q9Q1V4 human immun
59	7	4.1	132	Q53557	Q53557 salmonella
60	7	4.1	142	Q9SEW0	Q9SEW0 lilium long
61	7	4.1	147	Q95S08	Q95S08 drosophila
62	7	4.1	149	Q94102	Q94102 oryza sativ
63	7	4.1	149	Q9D7P2	Q9D7P2 mus musculu
64	7	4.1	150	Q9SA42	Q9SA42 arabidopsis
65	7	4.1	158	Q9Y195	Q9Y195 drosophila
66	7	4.1	158	Q9CTR3	Q9CTR3 mus musculu
67	7	4.1	163	Q94GR5	Q94GR5 oryza sativ
68	7	4.1	163	Q9KZF6	Q9KZF6 streptomyce
69	7	4.1	165	Q07218	Q07218 mycobacteri
70	7	4.1	168	Q51489	Q51489 pseudomonas
71	7	4.1	168	Q9NEE1	Q9NEE1 leishmania
72	7	4.1	168	Q914Z4	Q914Z4 pseudomonas
73	7	4.1	168	Q98P85	Q98P85 rhizobium 1
74	7	4.1	170	Q4662	Q4662 bruceella ab
75	7	4.1	170	Q8YMX7	Q8YMX7 anabaena sp
76	7	4.1	171	Q9VPI1	Q9VPI1 drosophila
77	7	4.1	172	Q8Y1F4	Q8Y1F4 ralsstonia s
78	7	4.1	173	Q922F9	Q922F9 mus musculu
79	7	4.1	174	Q91WU2	Q91WU2 oryza sativ
80	7	4.1	174	Q8XFM6	Q8XFM6 salmonella
81	7	4.1	174	Q9YBA7	Q9YBA7 aeropyrum p
82	7	4.1	176	Q9SMY4	Q9SMY4 arabidopsis
83	7	4.1	176	Q926C3	Q926C3 rhizobium m
84	7	4.1	177	Q8U9U5	Q8U9U5 agrobacteri
85	7	4.1	179	Q9RTG4	Q9RTG4 leishmania
86	7	4.1	180	Q9NMR9	Q9NMR9 leishmania
87	7	4.1	184	Q919S1	Q919S1 white spot
88	7	4.1	185	Q9PEZ5	Q9PEZ5 xylella fas
89	7	4.1	186	Q30783	Q30783 chlamydophi

90 7 4.1 188 16 084605 084605 chlamydia t
91 7 4.1 192 16 0927C5 0927C5 chlamydia p
92 7 4.1 196 5 09N7X4 09N7X4 leishmania
93 7 4.1 197 2 09X6J7 09X6J7 vibrio chol
94 7 4.1 198 10 09FTW8 09FTW8 oryza sativ
95 7 4.1 199 2 08VOT5 08VOT5 vibrio chol
96 7 4.1 199 2 08VOT4 08VOT4 vibrio chol
97 7 4.1 199 8 002667 002667 podospora a
98 7 4.1 200 10 09JN12 09JN12 arabidopsis
99 7 4.1 202 16 09P2F0 09P2F0 chlamydia m
100 7 4.1 206 10 0946V3 0946V3 zea mays (m

ALIGNMENTS

RESULT 1

Q9PC85 PRELIMINARY; PRT; 186 AA.
AC Q9PC85;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein Pe precursor.
GN XFI896.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Boye J.M., Bionnes M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorriy H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Klieger J.E., Kurama E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nant A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawaaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silvestri W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zaitz M., Weidman J., Setubal J.C.;
RT "the genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AEO04009; AAF84702.1; -;
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_OmpA; 1.
KW Complete proteome.
SQ SEQUENCE 186 AA; 20431 MW; C8DD2F6233DB0C92 CRC64;

Query Match 5.8%; Score 10; DB 16; Length 186;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 TDERGSRVYN 117
Db 119 TDERGSRVYN 128

RESULT 2

Q94EA6 PRELIMINARY; PRT; 998 AA.
AC Q94EA6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE P0435H01.25 protein.
GN P0435H01.25.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0435H01."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003142; BAB63529.1; -;
DR InterPro: IPR000862; RRCdomain.
SQ SEQUENCE 998 AA; 108459 MW; 0826071700585730 CRC64;

Query Match 5.2%; Score 9; DB 10; Length 998;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 16
Db 170 AAAAALSV 178

RESULT 3

Q9VOM0 PRELIMINARY; PRT; 2168 AA.
AC Q9VOM0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE TOC protein.
GN TOC OR C93660.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiot E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelfand W.M., Glasser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,
 RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
 RA SVIRSKAS R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang X., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003581; AAF51145.1; -;
 DR FLYBASE: FBgn0015600; loc.
 DR InterPro: IPR000169; SHPOT_acsite.
 DR PROSITE: PS00639; TH1OL_PROTEASE_HIS; UNKNOWN_1.
 SO SEQUENCE 2168 AA; 234786 MW; F26344DB2E1CAF4D CRC64;

Query Match 5.2%; Score 9; DB 5; Length 2168;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 AAAAAALSV 16
 Db 973 AAAAAALSV 987

RESULT 4
 ID 046112 PRELIMINARY; PRT; 2176 AA.
 AC 046112;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE TOUCAN protein.
 GN TOC OR CG9660.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ADULT;
 RX MEDLINE=98090047; PubMed=9362455;
 RA Grammont M., Dastugue B., Coudere J.L.;
 RT "The Drosophila toucan (toc) gene is required in germline cells for
 RL the somatic cell patterning during oogenesis.";
 RL Development 124:4917-4926(1997).
 DR EMBL: Y14157; CAA74574.1; -;
 DR FLYBASE: FBgn0015600; loc.
 DR InterPro: IPR000169; SHPOT_acsite.
 DR PROSITE: PS00639; TH1OL_PROTEASE_HIS; UNKNOWN_1.
 SO SEQUENCE 2176 AA; 235405 MW; 35ABDBE00B49EFC7 CRC64;

Query Match 5.2%; Score 9; DB 5; Length 2176;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 AAAAAALSV 16
 Db 973 AAAAAALSV 981

RESULT 5
 ID 09X0G9 PRELIMINARY; PRT; 143 AA.
 AC 09X0G9;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sugar-phosphate isomerase.
 GN TM1080.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571.
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RL genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).
 DR EMBL: AE001768; AAD36157.1; -;
 DR TIGR: TM1080; -;
 DR InterPro: IPR003500; Rib/Gal_isomerase.
 DR Pfam: PF02502; LACAB_r1B; 1.
 DR TIGRFAMs: TIGR00689; r1B_LacA_LacB; 1.
 KM Isomerase; Complete proteome.
 SO SEQUENCE 143 AA; 13867 MW; 3E17D2A10FCCBD CRC64;

Query Match 4.7%; Score 8; DB 16; Length 143;
 Best Local Similarity 100.0%; Pred. No. 7.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 NYLKGKI 137
 Db 20 NYLKGKI 27

RESULT 6
 ID 08S1R2 PRELIMINARY; PRT; 249 AA.
 AC 08S1R2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative UDP-glucuronic acid decarboxylase.
 GN P0506H12.31.
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Erihartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RL clone: P0506H12.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBD databases.
 DR EMBL: AP003271; BAB89759.1; -;
 SO SEQUENCE 249 AA; 26592 MW; 7954F613B241E81F CRC64;

Query Match 4.7%; Score 8; DB 10; Length 249;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 SKIPSIYV 71
 Db 11111111

Db 25 SKIPSLVY 32

RESULT 7

09B1J7 PRELIMINARY; PRT; 258 AA.

AC 09B1J7; 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE DNA-binding transcriptional repressor giant (Fragment).

GN GR.
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;

RP [1]
RA SEQUENCE FROM N.A.
RA Thomas J., Arnosti D.N.,
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF356543; AAK28631.1; -
DR FlyBase; FBgn0044278; Dhyd\gtf.

FT DNA-binding. 1
FT NON_TER 258 1
FT SEQUENCE 258 AA; 26948 MW; 09C0CA545CE455EA CRC64;

Query Match 4.7%; Score 8; DB 5; Length 258;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALS 15
Db 31 AAAAALS 38

RESULT 8

093NP5 PRELIMINARY; PRT; 268 AA.

AC 093NP5; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Wbwc.

GN Wbwc.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;

RP [1]
RA SEQUENCE FROM N.A.
RA MEDLINE-21297197; PubMed-11404020;
RA Wang L., Briggs C.E., Rothmund D., Fratamico P., Luchansky J.B.,
RA Reeves P.R.;
RT Sequence of the E. coli O104 antigen gene cluster and identification
RT of O104 specific genes.";
RL Gene 270:231-236(2001).

DR EMBL; AF361371; AAK64375.1; -
DR InterPro; IPR001173; Glycos.transf_2.
DR Pfam; pf00535; Glycos.transf_2; 1.
SQ SEQUENCE 268 AA; 31176 MW; C1EF30C136A668F7 CRC64;

Query Match 4.7%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 TGVITNGV 49
Db 239 TGVITNGV 246

RESULT 9

0930C4 PRELIMINARY; PRT; 280 AA.

AC 0930C4; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein RA0274.

GN RA0274 OR SMA0520.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;

RP [1]
RA SEQUENCE FROM N.A.
RA SPRAIN-1021;
RA MEDLINE-21396509; PubMed-11481432;

RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barlow-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Hultzer L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kallman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSym megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).

DR EMBL; AE007220; AAK64932.1; -
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 280 AA; 30615 MW; FF533F65700FC352 CRC64;

Query Match 4.7%; Score 8; DB 16; Length 280;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALS 15
Db 121 AAAAALS 128

RESULT 10

09VVI8 PRELIMINARY; PRT; 306 AA.

AC 09VVI8; 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE CG6273 protein.

GN CG6273.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

RP [1]
RA SEQUENCE FROM N.A.
RA STRAIN-BERKELEY;
RA MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer G., Nelson C.R., Baldwin D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Andrews-Pfannkoch C.,
RA Abril J.F., Agbayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
RA Balow R.M., Basu A., Baxendale J., Bhandari D., Bolshakov S.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
RA Borovoy D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlke C., Mays A.D., Dew I., Dietz S.M.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dunn B.C., Dunn P.,
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Fierstein S.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mount S.M., Moy M., Murphy B., Murphy C., Morris J., Moshrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Nelson D.L.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AEO03523; AAF49323.1; -
DR FlyBase: FBgn0036739; CG6273.
SQ SEQUENCE 306 AA; 30932 MW; 8C26CE0BEDDADA3 CRC64;

Query Match 4.7%; Score 8; DB 5; Length 306;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAAAAAAL 14
Db 195 TAAAAAAL 202

RESULT 11
O986T0 PRELIMINARY; PRT; 316 AA.
AC O986T0:
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein mlr7223.
GN MLR7223.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Matsumoto A., Iida S., Ishikawa A., Kawasumi K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003011; BAB53373.1; -
DR InterPro: IPR002173; Pfam.
DR Pfam: PF00294; Pfam: 1.
DR PROSITE: PS00584; PFAM_KINASES_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 316 AA; 32961 MW; 258AE2AVBE89DFID CRC64;

Query Match 4.7%; Score 8; DB 16; Length 316;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
Db 269 AAAAAAALS 276

RESULT 12
O9NGI9 PRELIMINARY; PRT; 324 AA.
AC O9NGI9:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Eye1d (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SIM3;
RX MEDLINE=20283933; PubMed=10823947;
RA Begun D.J., Whitely P.;
RT "Reduced X-linked nucleotide polymorphism in *Drosophila simulans*.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL: AF252666; AAF68048.1; -
DR FlyBase: FBgn0041660; Dsim\osa.
FT NON_TER 1 324 1
FT NON_TER 324 324 1
SQ SEQUENCE 324 AA; 32437 MW; 446774235A5071F0 CRC64;

Query Match 4.7%; Score 8; DB 5; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
Db 269 AAAAAAALS 276

RESULT 13
O9NGI8 PRELIMINARY; PRT; 324 AA.
AC O9NGI8:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Eye1d (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SIM4;
RX MEDLINE=20283933; PubMed=10823947;
RA Begun D.J., Whitely P.;
RT "Reduced X-linked nucleotide polymorphism in *Drosophila simulans*.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL: AF252667; AAF68049.1; -
DR FlyBase: FBgn0041660; Dsim\osa.
FT NON_TER 1 324 1
FT NON_TER 324 324 1
SQ SEQUENCE 324 AA; 32373 MW; 5EC6A8C4546A230 CRC64;

Query Match 4.7%; Score 8; DB 5; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
Db 269 AAAAAAALS 276

```
RESULT 14
O9NG17 PRELIMINARY; PRT: 324 AA.
ID O9NG17:
AC 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Eyelid (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SIM5;
RX MEDLINE=20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF252668; AAF68050.1; -.
DR Flybase; FBgn0041660; Dsim\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32399 MW; 5ECC6A8C4556B331 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 324;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
DB 269 AAAAAALS 276

RESULT 15
O9NGB4 PRELIMINARY; PRT: 324 AA.
ID O9NGB4:
AC 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Eyelid (Fragment).
GN OSA.
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF255314; AAF68611.1; -.
DR Flybase; FBgn0041640; Dyak\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32644 MW; 65BA271ED4DE03D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 324;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
DB 269 AAAAAALS 276

RESULT 16
```

```
O9NGK3 PRELIMINARY; PRT: 324 AA.
ID O9NGK3:
AC 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Eyelid (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SIM7, AND SIM2;
RX MEDLINE=20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF252670; AAF68052.1; -.
DR Flybase; FBgn0041660; Dsim\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32403 MW; 4461ADE6E896B334 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 324;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
DB 269 AAAAAALS 276

RESULT 17
O9NGK2 PRELIMINARY; PRT: 324 AA.
ID O9NGK2:
AC 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Eyelid (Fragment).
GN OSA.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SIM8, AND SIM6;
RX MEDLINE=20283933; PubMed=10823947;
RA Begun D.J., Whitley P.;
RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
DR EMBL; AF252671; AAF68053.1; -.
DR Flybase; FBgn0041660; Dsim\osa.
FT NON_TER 1 324
SQ SEQUENCE 324 AA; 32407 MW; 44676F22EA5071F0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 324;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
DB 269 AAAAAALS 276
```

RESULT 18

08X1X2 PRELIMINARY; PRT; 381 AA.
 AC 08X1X2;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Pectin lyase A.
 GN PEIA.
 OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 CC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 NX NCBI_TaxID=5062;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KBN616;
 RA Kitamoto N.;
 RT "Two pectin lyase genes (peia and pelb) from Aspergillus oryzae
 RT KBN616: their sequence analyses and overexpression, and
 RT characterization of the gene products."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB029322; BAB82467.1;
 DR InterPro: IPR002022; Amb_allergen.
 DR Pfam: PF00544; pec_lyase; 1.
 KW Lyase.
 SQ SEQUENCE 381 AA; 39982 MW; BA6FC139D31E10F7 CRC64;

Query Match 4.7%; Score 8; DB 16; Length 381;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAAAL 14
 DB 7 IAAAAAAL 14

RESULT 19
 09J5Z1 PRELIMINARY; PRT; 452 AA.
 AC 09J5Z1;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE UDP-MurNAC-pentapeptide synthetase (Ec 6.3.2.15).
 GN MURF OR NMA2068.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=63699;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-22491 / SEROGROUP A / SPROTYPE 4A.
 RX MEDLINE-20222556; PubMed-10761919;
 RA Parthill J., Achman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 RA Jags K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 RT meningitidis 22491."
 RL Nature 404:502-506(2000).
 DR EMBL; AL162758; CAB85286.1;
 DR HSSP; P11880; 1G64.
 DR InterPro: IPR000713; Mur_lyase.
 DR InterPro: IPR004101; Mur_lyase_C.
 DR Pfam: PF01225; Mur_lyase; 1.
 DR Pfam: PF02875; Mur_lyase_C; 1.
 DR TIGRfams; TIGR01143; murf; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 452 AA; 48112 MW; 9903D4DD1B64CA9 CRC64;

Query Match 4.7%; Score 8; DB 16; Length 452;
 Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAAL 15
 DB 289 AAAAAAAL 296

RESULT 20

096IK5 PRELIMINARY; PRT; 504 AA.
 AC 096IK5;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein mlr2359.
 GN MLR2359.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Phyllobacteriaceae; Mesorhizobium.
 NX NCBI_TaxID=381;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFP303099;
 RX MEDLINE-21082930; PubMed-11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno M.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP002999; BAB49511.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 504 AA; 53754 MW; ABDCC1D5759DD0DB CRC64;

Query Match 4.7%; Score 8; DB 16; Length 504;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAAAL 14
 DB 282 IAAAAAAL 289

RESULT 21

09HTN7 PRELIMINARY; PRT; 526 AA.
 AC 09HTN7;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Probable binding protein component of ABC dipeptide transporter.
 GN PA5317.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 OX NCBI_TaxID=287;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE-20437337; PubMed-10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
 RT opportunistic pathogen."

RL Nature 406:959-964(2000).
 EMBL: AE004944; AAG08702.1; --
 DR HSSP: P23847; 1DPE.
 DR InterPro: IPR000914; SBP_bac_5.
 DR Pfam: PF00496; SBP_bac_5; 1.
 DR Complete proteome.
 SW SEQUENCE 526 AA; 57880 MW; EE3F4B3A0914B24B CRC64;

Query Match 4.7%; Score 8; DB 16; Length 526;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AAAAALSV 16
 Db 19 AAAAALSV 26

RESULT 22

ID 0957V5 PRELIMINARY; PRT: 541 AA.
 AC 0957V5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE T16011.4 protein (AF360900/716011_4).
 GN T16011.4 OR MZB10.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronging C.M., Koo H., Fujil C.Y., Uterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III BAC T16011 genomic sequence."; Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronging C.M., Koo H., Fujil C.Y., Uterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III PI MZB10 genomic sequence."; Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RA "Arabidopsis cDNA clones";
 RT Ecker J.R.;
 RA "Arabidopsis cDNA clones";
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC010871; AAF07827.1; --
 DR EMBL: AC009326; AAD56316.1; --
 DR EMBL: AF428276; AAL16108.1; --

SO SEQUENCE 541 AA; 56813 MW; 3AEFE15AD4EAD6E CRC64;

Query Match 4.7%; Score 8; DB 10; Length 541;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 15
 Db 43 AAAAALSV 50

RESULT 23

ID 095S11 PRELIMINARY; PRT: 545 AA.
 AC 095S11;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE H102140P.
 GN PUBSF OR CG12085.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Broksstein P., Hong L., Aghayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Lao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Ceiliker S.;
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A1061006; AAL28554.1; --
 DR Flybase: FBgn028577; pubsf.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 3.
 DR PROSITE: PS50102; RRM; 3.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_2.
 SO SEQUENCE 545 AA; 57989 MW; E75E378FD07E24E CRC64;

Query Match 4.7%; Score 8; DB 5; Length 545;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALSV 15
 Db 369 AAAAALSV 376

RESULT 24

ID 09X4X4 PRELIMINARY; PRT: 547 AA.
 AC 09X4X4;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE DITE.
 GN DITE.
 OS Pseudomonas abietaniphila.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas;
 OX NCBI_TaxID=89065;

RP SEQUENCE FROM N.A.
 RC STRAIN=BKME-9;
 RA Mohn W.W., Wilson A.E., Bicho P., Moore E.R.B.;
 RT "Physiological and Phylogenetic Diversity of Bacteria Growing on Resin
 Acids";
 RT Acids; Appl. Microbiol. 0:0-0(1999).
 RL [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BKME-9;
 RA MEDLINE=99235742; PubMed=10217753;
 RA Martin V.J., Mohn W.W.;
 RT "A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid-
 degrading bacterium Pseudomonas abietaniphila BKME-9";
 RT J. Bacteriol. 181:2675-2682(1999).
 DR EMBL: AF119621; AAD21067.1; --

SO SEQUENCE 547 AA; 58500 MW; 49A627265D7C1A74 CRC64;

Query Match 4.7%; Score 8; DB 2; Length 547;
 Best Local Similarity 100.0%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AAAALSVL 17
 |||||
 DB 107 AAAALSVL 114

RESULT 25

O9M7S3 PRELIMINARY; PRT; 570 AA.

AC O9M7S3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 4-comurate-coa ligase 4CCL1 (EC 6.2.1.12).
 OS Lollum perenne (Perennial ryegrass).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Poace; Lollum.
 OX NCBI_Taxid=4522;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. ELLET;
 RA Heath R.L., Huxley H., Spangenberg G.;
 RT "Isolation of three 4-comurate-coa ligase cDNA homologues from
 Lollum perenne".
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF052221; AAF3732.1; -.
 DR HSSP; P08659; 1LCI.
 DR InterPro; IPR000873; AMP-Bind.
 DR Pfam; PF00501; AMP-binding; 1.
 DR PROSITE; PS00455; AMP_BINDING; 1.
 KW Ligase.
 SQ SEQUENCE 570 AA; 60293 MW; 8194E1B939D2925 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 8; DB 10; Length 570;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 QIAAAAAA 13
 |||||
 DB 13 QIAAAAAA 20

RESULT 26

O8VOPT PRELIMINARY; PRT; 587 AA.

AC O8VOPT;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Flagellar MS ring protein.
 GN FLIF.
 OS Burkholderia cepacia (Pseudomonas cepacia).
 OC Bacteria; Proteobacteria; Beta subdivision; Burkholderia group;
 OC Burkholderia.
 OX NCBI_Taxid=292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-J2315;
 RA Tomich M., Herfst C.A., Golden J.W., Mohr C.D.;
 RT "Role of Flagella in Burkholderia cepacia Host Cell Invasion."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF453480; AAL65160.1; -.
 DR InterPro; IPR000067; FlgMing_FLIF.
 DR InterPro; IPR002920; YscJ_FLIF.
 DR Pfam; PF01514; YscJ_FLIF; 1.
 DR PRINTS; PR01009; FLGMRINGFLIF.
 DR TIGRfams; TIGR00206; flif; 1.
 SQ SEQUENCE 587 AA; 62991 MW; 4B142D53E065CB91 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 8; DB 2; Length 587;
 Matches 100.0%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 IAAAAAL 14
 |||||
 DB 491 IAAAAAL 498

RESULT 27

O02427 PRELIMINARY; PRT; 587 AA.

AC O02427;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Forkhead homolog.
 GN CF-FKH.
 OS Clona intestinalis.
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 OC Phlebobranchia; Clonidae; Clona.
 OX NCBI_Taxid=7719;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Corbo J.C., Erives A., Di Gregorio A., Chang A., Levine M.;
 RT "Horseshoe patterning of the vertebrate neural tube is conserved in
 a protochordate."
 RL Development 0:0-0(1997).
 DR EMBL; AF002988; AAB61227.1; -.
 DR HSSP; Q63245; 2HFH.
 DR InterPro; IPR001766; TF_Fork_head.
 DR Pfam; PF00250; Fork_head; 1.
 DR PRINTS; PR00053; FORKHEAD.
 DR Prodom; PD000425; TF_Fork_head; 1.
 DR SMART; SM00339; FH; 1.
 DR PROSITE; PS00657; FORK_HEAD_1; 1.
 DR PROSITE; PS00658; FORK_HEAD_2; 1.
 DR PROSITE; PS00039; FORK_HEAD_3; 1.
 SQ SEQUENCE 587 AA; 64092 MW; 08051624221A9D00 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 8; DB 5; Length 587;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 QIAAAAAA 13
 |||||
 DB 80 QIAAAAAA 87

RESULT 28

O9W2T0 PRELIMINARY; PRT; 624 AA.

AC O9W2T0;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE CG15295 protein.
 GN CG15295.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelhiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mitos G.L.G.,
 RA April J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Foster C., Gabriellian A.E., Garq N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jatali M., Kalush F., Karpem G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Pittman G.S., Pan S., Scheeler F., Shen H.,
 RA Reinert K., Remington K., Saunders R.D.C., Skupski M.P., Smith T.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AF003451; AAF46509.1; -
 DR Flybase: FBgn0030220; CG15295.
 DR FLYBASE 624 AA; 65843 MW; 8960041374AFD6C1 CRC64;
 SQ

Query Match 4.7%; Score 8; DB 5; Length 624;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
 |||||||
 Db 158 AAAAAALS 165

RESULT 29
 OQ9W0E6 PRELIMINARY; PRT; 637 AA.
 ID OQ9W0E6; OQ9W0E7; OQ9W0E6; 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Poly-U-binding-splicing-factor protein (CG12085 protein).
 GN PUSF OR POLY-U-BINDING-SPlicing-FACTOR OR CG12085.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aydayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu U., Besley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garq N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jatali M., Kalush F., Karpem G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Pittman G.S., Pan S., Scheeler F., Shen H.,
 RA Reinert K., Remington K., Saunders R.D.C., Skupski M.P., Smith T.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AF479079; AAL86452.1; -
 DR FLYBASE: FBgn0030220; CG15295.
 DR FLYBASE 624 AA; 65843 MW; 8960041374AFD6C1 CRC64;
 SQ

Query Match 4.7%; Score 8; DB 5; Length 637;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAALS 15
 |||||||
 Db 461 AAAAAALS 468

RESULT 30
 O8T6B9 PRELIMINARY; PRT; 637 AA.
 ID O8T6B9; O8T6B9; 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Half pint.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Van Buskirk C., Schupbach T.;
 RA "Half pint regulates alternative splice site selection in
 RT *Drosophila*.";
 RL Dev. Cell 0:0-0(2002).
 DR EMBL: AF479079; AAL86452.1; -

SO SEQUENCE 637 AA; 67895 MW; 00A81D4D9EF1282 CRC64;
 Query Match 4.7%; Score 8; DB 5; Length 637;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||
 Db 461 AAAAAAALS 468

RESULT 31

O9FRV3 PRELIMINARY; PRT; 644 AA.
 AC O9FRV3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Acetolactate synthase.
 GN ALS.
 OS Oryza sativa (Rice)
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eumariophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriocaridaceae; Oryzaeae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV.KINMAZE;
 RA Shimizu T., Kato Y., Nakayama I., Nakayama K., Fukuda A., Tanaka Y.;
 RT "Isolation and Expression of acetolactate synthase genes from Oryza
 sativa.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
 DR EMBL: AB049822; BAB20812.1;
 DR InterPro: IPR004407; AcoIac_19.
 DR InterPro: IPR000399; TPP_enzyme.
 DR Pfam: PF00205; TPP_enzymes; 1.
 DR Pfam: PF02775; TPP_enzymes; C; 1.
 DR Pfam: PF02776; TPP_enzymes; N; 1.
 DR TIGRfams: TIGR00118; acoIac_19; 1.
 DR KEGG: K0118; acoIac_19; 1.
 KW Flavoprotein; Lyase; Thiamine pyrophosphate.
 SO SEQUENCE 644 AA; 69392 MW; EDB26E5F8B89EC9 CRC64;

Query Match 4.7%; Score 8; DB 10; Length 644;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||
 Db 6 AAAAAAALS 13

RESULT 32

O9FRV2 PRELIMINARY; PRT; 644 AA.
 AC O9FRV2;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Acetolactate synthase.
 GN ALS.
 OS Oryza sativa (Rice)
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Eumariophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriocaridaceae; Oryzaeae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV.KINMAZE;
 RA Shimizu T., Kato Y., Nakayama I., Nakayama K., Fukuda A., Tanaka Y.;
 RT "Isolation and Expression of acetolactate synthase genes from Oryza
 sativa.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

CC -1 SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
 DR EMBL: AB049823; BAB20813.1;
 DR InterPro: IPR004407; AcoIac_19.
 DR InterPro: IPR000399; TPP_enzyme.
 DR Pfam: PF00205; TPP_enzymes; 1.
 DR Pfam: PF02775; TPP_enzymes; C; 1.
 DR Pfam: PF02776; TPP_enzymes; N; 1.
 DR TIGRfams: TIGR00118; acoIac_19; 1.
 DR KEGG: K0118; acoIac_19; 1.
 KW Flavoprotein; Lyase; Thiamine pyrophosphate.
 SO SEQUENCE 644 AA; 69345 MW; F873F121A9FD9F99 CRC64;

Query Match 4.7%; Score 8; DB 10; Length 644;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAAALS 15
 |||||
 Db 6 AAAAAAALS 13

RESULT 33

O9HKB4 PRELIMINARY; PRT; 651 AA.
 AC O9HKB4;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE Hypothetical protein Ta0687.
 GN TA0687.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Repp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermophilic scavenger Thermoplasma
 acidophilum.";
 RL Nature 407:508-513 (2000).
 DR EMBL: AL445065; CAC11825.1;
 DR Hypothetical protein; Complete proteome.
 SO SEQUENCE 651 AA; 74410 MW; 3F9E9F1711E7B733 CRC64;

Query Match 4.7%; Score 8; DB 17; Length 651;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 LHIQIATAA 10
 |||||
 Db 128 LHIQIATAA 135

RESULT 34

O9V9F7 PRELIMINARY; PRT; 662 AA.
 AC O9V9F7;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CG10417 protein (LD27655P).
 GN CG10417.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;


```

O9BMZ8
ID O9BMZ8 PRELIMINARY; PRT; 825 AA.
AC O9BMZ8;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DE 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE AF10.
GN ALHAMBRA OR CG1070.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Linder B., Jackle H.;
RT "The Drosophila homolog of AF10 is expressed in a striped pattern
RL throughout blastoderm."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF281146; AAK06386.1; -
DR FlyBase; FBgn0037471; Alhambra.
SQ SEQUENCE 825 AA; 84483 MW; 9A2C7A795B9F43D6 CRC64;

Query Match
Best Local Similarity 4.7%; Score 8; DB 5; Length 825;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALSL 15
DB 601 AAAAALSL 608

RESULT 38
O962D2 PRELIMINARY; PRT; 975 AA.
AC O962D2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DE 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Zinc finger/leucine zipper protein DALL isoform C1.
GN ALHAMBRA OR CG1070.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21095180; PubMed=11165485;
RA Bahri S.M., Chia W., Yang X.;
RT "The Drosophila homolog of human AF10/AF17 leukemia fusion genes
RT (Dall) encodes a zinc finger/leucine zipper nuclear protein regulated
RT growth."
RL Mech. Dev. 100:291-301(2001).
DR EMBL; AY036104; AAK63170.1; -
DR FlyBase; FBgn0037471; Alhambra.
SQ SEQUENCE 975 AA; 98713 MW; 5BB70FB144935276 CRC64;

Query Match
Best Local Similarity 4.7%; Score 8; DB 5; Length 975;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALSL 15
DB 600 AAAAALSL 607

RESULT 39
O8SZM9 PRELIMINARY; PRT; 1015 AA.
AC O8SZM9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)

```

```

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE LP05501P.
GN ALHAMBRA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Paele J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY069737; AAL39882.1; -
SQ SEQUENCE 1015 AA; 100464 MW; CB24616CF13D4C53 CRC64;

Query Match
Best Local Similarity 4.7%; Score 8; DB 5; Length 1015;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAALSL 15
DB 791 AAAAALSL 798

RESULT 40
O9YI61 PRELIMINARY; PRT; 1164 AA.
AC O9YI61;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DE 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CG1070 protein.
GN ALHAMBRA OR CG1070.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Mortman J.R., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Zhang O., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burlis K.C., Busam M.R., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisi M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Rupp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Koshnefti A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massarman D.A., Weinstock G.M., Weissendbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gabbas R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
DR EMBL: AF003672; AAF54065.1; -
DR FlyBase: FBgn0037471; Alhambra.
SQ SEQUENCE 1164 AA; 119469 MW; F49C6BC7394DBB6A CRC64;

Query Match 4.7%; Score 8; DB 5; Length 1164;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAALS 15
|||||||
DB 940 AAAAALS 947

Search completed: July 6, 2003, 14:21:32
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:20:50 ; Search time 19 Seconds

(without alignments)
266.355 Million cell updates/sec

Title: US-09-674-779B-2

Perfect score: 172

Sequence: 1 MMLHQLIAAAALSLVTFM.....IAFTNEAMQNRRAELSY 172

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	4.1	9	US-08-159-339A-154	Sequence 154, App
2	7	4.1	13	PCT-US95-04121-52	Sequence 52, App
3	7	4.1	14	US-08-503-226B-42	Sequence 42, App
4	7	4.1	14	US-08-721-458B-42	Sequence 42, App
5	7	4.1	22	US-08-516-859A-103	Sequence 103, App
6	7	4.1	22	US-09-586-472-103	Sequence 103, App
7	7	4.1	22	US-09-528-706-103	Sequence 103, App
8	7	4.1	37	US-08-180-524-1	Sequence 1, App
9	7	4.1	37	US-08-180-524-8	Sequence 8, App
10	7	4.1	37	US-08-180-524-9	Sequence 9, App
11	7	4.1	37	US-08-975-166-1	Sequence 1, App
12	7	4.1	37	US-08-975-166-8	Sequence 8, App
13	7	4.1	37	US-08-975-166-9	Sequence 9, App
14	7	4.1	37	US-09-117-121-39	Sequence 39, App
15	7	4.1	37	US-09-117-121-42	Sequence 42, App
16	7	4.1	37	US-09-344-529-4	Sequence 4, App
17	7	4.1	38	US-07-814-220-1	Sequence 1, App
18	7	4.1	109	US-07-812-421-1	Sequence 1, App
19	7	4.1	109	US-09-325-932A-144	Sequence 144, App
20	7	4.1	203	US-08-233-609-3	Sequence 3, App
21	7	4.1	203	US-08-444-083-3	Sequence 3, App
22	7	4.1	203	US-08-286-304-3	Sequence 3, App
23	7	4.1	203	US-08-442-745-3	Sequence 3, App
24	7	4.1	203	US-08-443-129-3	Sequence 3, App
25	7	4.1	203	US-08-443-952-3	Sequence 3, App
26	7	4.1	203	US-08-443-130-3	Sequence 3, App
27	7	4.1	203	US-08-898-911-3	Sequence 3, App

28	7	4.1	203	PCT-US95-04467-3	Sequence 3, App
29	7	4.1	228	US-09-286-690-12	Sequence 12, App
30	7	4.1	309	US-08-508-761B-32	Sequence 32, App
31	7	4.1	325	US-08-107-676-30	Sequence 30, App
32	7	4.1	350	US-09-161-241-9	Sequence 9, App
33	7	4.1	482	US-07-792-885A-1	Sequence 1, App
34	7	4.1	482	US-08-142-439A-7	Sequence 7, App
35	7	4.1	482	US-08-869-477-7	Sequence 7, App
36	7	4.1	503	US-09-172-952-26	Sequence 26, App
37	7	4.1	503	US-09-382-286-10	Sequence 10, App
38	7	4.1	503	US-09-395-115-10	Sequence 10, App
39	7	4.1	503	US-08-436-285-10	Sequence 10, App
40	7	4.1	503	US-09-679-187-10	Sequence 10, App
41	7	4.1	674	US-08-893-852A-1	Sequence 1, App
42	7	4.1	760	US-08-195-152-2	Sequence 2, App
43	7	4.1	894	US-09-735-934A-4	Sequence 4, App
44	7	4.1	1084	US-09-237-725A-3	Sequence 3, App
45	7	4.1	1724	US-08-857-076-12	Sequence 12, App
46	7	4.1	7257	US-09-335-409-5	Sequence 5, App
47	7	4.1	7257	US-09-568-102-5	Sequence 5, App
48	7	4.1	7257	US-09-567-969-5	Sequence 5, App
49	7	4.1	7257	US-09-568-480-5	Sequence 5, App
50	7	4.1	7257	US-09-568-486-5	Sequence 5, App
51	7	4.1	7257	US-09-568-472-5	Sequence 5, App
52	7	4.1	7257	US-09-567-899-5	Sequence 5, App
53	6	3.5	6	US-08-837-305-1	Sequence 1, App
54	6	3.5	6	US-08-611-395-9	Sequence 9, App
55	6	3.5	7	US-09-122-126B-5	Sequence 5, App
56	6	3.5	8	US-08-963-168C-24	Sequence 24, App
57	6	3.5	8	US-08-963-168C-26	Sequence 26, App
58	6	3.5	9	US-08-425-069-12	Sequence 12, App
59	6	3.5	9	US-08-317-844B-12	Sequence 12, App
60	6	3.5	9	US-08-396-385-12	Sequence 12, App
61	6	3.5	9	US-08-159-339A-129	Sequence 129, App
62	6	3.5	9	US-08-159-339A-143	Sequence 143, App
63	6	3.5	9	US-08-159-339A-151	Sequence 151, App
64	6	3.5	9	US-08-159-339A-152	Sequence 152, App
65	6	3.5	9	US-08-159-339A-153	Sequence 153, App
66	6	3.5	9	US-08-159-339A-155	Sequence 155, App
67	6	3.5	9	US-08-159-339A-156	Sequence 156, App
68	6	3.5	9	US-08-159-339A-157	Sequence 157, App
69	6	3.5	9	US-08-159-339A-158	Sequence 158, App
70	6	3.5	9	US-08-159-339A-159	Sequence 159, App
71	6	3.5	9	US-08-159-339A-160	Sequence 160, App
72	6	3.5	9	US-08-159-339A-161	Sequence 161, App
73	6	3.5	9	US-08-159-339A-162	Sequence 162, App
74	6	3.5	9	US-08-159-339A-163	Sequence 163, App
75	6	3.5	9	US-08-159-339A-164	Sequence 164, App
76	6	3.5	9	US-08-159-339A-165	Sequence 165, App
77	6	3.5	9	US-08-159-339A-166	Sequence 166, App
78	6	3.5	9	US-08-159-339A-167	Sequence 167, App
79	6	3.5	9	US-08-159-339A-168	Sequence 168, App
80	6	3.5	9	US-08-159-339A-169	Sequence 169, App
81	6	3.5	9	US-08-159-339A-170	Sequence 170, App
82	6	3.5	9	US-09-287-221-12	Sequence 12, App
83	6	3.5	10	US-08-406-193-14	Sequence 14, App
84	6	3.5	10	US-08-406-193-15	Sequence 15, App
85	6	3.5	10	US-08-463-863-9	Sequence 9, App
86	6	3.5	10	US-08-545-151-14	Sequence 14, App
87	6	3.5	10	US-08-545-151-15	Sequence 15, App
88	6	3.5	10	US-08-458-887-9	Sequence 9, App
89	6	3.5	10	US-08-159-339A-149	Sequence 149, App
90	6	3.5	10	US-08-159-339A-150	Sequence 150, App
91	6	3.5	10	US-08-963-168C-25	Sequence 25, App
92	6	3.5	10	5169933-18	Sequence 18, App
93	6	3.5	10	5169933-19	Sequence 19, App
94	6	3.5	10	5169933-41	Sequence 41, App
95	6	3.5	10	US-07-814-220-5	Sequence 5, App
96	6	3.5	11	US-07-814-220-6	Sequence 6, App
97	6	3.5	11	US-07-814-220-8	Sequence 8, App
98	6	3.5	11	US-07-814-220-11	Sequence 11, App
99	6	3.5	11	US-07-812-421-5	Sequence 5, App
100	6	3.5	11		

ALIGNMENTS

RESULT 1
US-08-159-339A-154
Sequence 154, Application US/08159339A
Patent No. 6037135
GENERAL INFORMATION:
APPLICANT: Kubo, Ralph T.
APPLICANT: Grey, Howard M.
APPLICANT: Settle, Alessandro
APPLICANT: Settle, Esteban
TITLE OF INVENTION: HLA Binding peptides and their
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 1254
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,339A
FILING DATE: 29-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,666
FILING DATE: 07-AUG-1992
APPLICATION NUMBER: US 08/027,746
FILING DATE: 05-MAR-1993
APPLICATION NUMBER: US 08/103,396
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-005030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX:
INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-154

Query Match 4.1%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAAAAA 13
|||||
DB 2 TAAAAA 8

RESULT 2
PCT-US95-04121-52
Sequence 52, Application PC/TUS9504121
GENERAL INFORMATION:
APPLICANT:
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: Haptened peptides and Uses Thereof
NUMBER OF SEQUENCES: 62

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04121
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/222,206
FILING DATE: April 1, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Vanstone, Darlene A.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 079.2PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6010
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
PCT-US95-04121-52

Query Match 4.1%; Score 7; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAAAAA 13
|||||
DB 3 TAAAAA 9

RESULT 3
US-08-503-226B-42
Sequence 42, Application US/08503226B
Patent No. 5871945
GENERAL INFORMATION:
APPLICANT: Lockertle, Robert Owen, et al.
TITLE OF INVENTION: Calcineurin Inhibitory Compounds and
TITLE OF INVENTION: Anchoring Protein
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/503,226B
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/404,731
FILING DATE: 15-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,227
FILING DATE: 23-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32861

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-6300
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-503-226B-42

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
DB 5 AAAAAA 11

RESULT 4
US-08-721-458B-42
Sequence 42, Application US/08721458B
Patent No. 6107104
GENERAL INFORMATION:
APPLICANT: Lockebble, Robert Owen, et al.
TITLE OF INVENTION: Calcineurin Inhibitory Compounds
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721.458B
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/404,731
FILING DATE: 15-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,227
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/503,226
FILING DATE: 17-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33276
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-6300
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-721-458B-42

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 14;
Matches 100.0%; Pred. No. 2.3;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 AAAAAA 14
DB 5 AAAAAA 11

RESULT 5
US-08-516-859A-103
Sequence 103, Application US/08516859A
Patent No. 6069231
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/516,859A
FILING DATE: 18-AUG-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-516-859A-103

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
DB 6 AAAAAA 12

RESULT 6
US-09-586-472-103
Sequence 103, Application US/09586472
Patent No. 632335
GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/586,472
FILING DATE: 01-Jun-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/528,706
FILING DATE: 17-MAR-2000
APPLICATION NUMBER: US 08/516,859
FILING DATE: 18-AUG-1995
APPLICATION NUMBER: US 08/399,411
FILING DATE: 06-MAR-1995
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 4130
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-09-586-472-103
SEQUENCE DESCRIPTION: SEQ ID NO: 103:

Query Match 4.1%; Score 7; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
|||||
DB 6 AAAAAA 12

RESULT 7
US-09-528-706-103
Sequence 103, Application US/09528706
Patent No. 6468985

GENERAL INFORMATION:
APPLICANT: Huang, Shi
TITLE OF INVENTION: Retinoblastoma Protein - Interacting
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/528,706
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/516,859

FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/292,683
FILING DATE: 18-AUG-1994

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1776
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: linear

US-09-528-706-103

Query Match 4.1%; Score 7; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
|||||
DB 6 AAAAAA 12

RESULT 8
US-08-180-524-1
Sequence 1, Application US/08180524
Patent No. 5849337

GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chitoye, Elzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudopleuronectes americanus
US-08-180-524-1

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
DB 6 AAAAAA 12

RESULT 9
US-08-180-524-8
Sequence 8, Application US/08180524
Patent No. 5849537
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kol, Edward
APPLICANT: Chicoye, Elzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thad Kryshak, Charles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-180-524-8

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
DB 6 AAAAAA 12

RESULT 10
US-08-180-524-9
Sequence 9, Application US/08180524
Patent No. 5849537
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kol, Edward
APPLICANT: Chicoye, Elzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thad Kryshak, Charles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-180-524-9

Query Match 4.1%; Score 7; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
|111111|
Db 6 AAAAAL 12

RESULT 11
US-08-975-166-1
Sequence 1, Application US/08975166
Patent No. 5928877
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,166
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO

ORIGINAL SOURCE:
ORGANISM: Pseudopleuronectes americanus
US-08-975-166-1

Query Match 4.1%; Score 7; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
|111111|
Db 6 AAAAAL 12

RESULT 12
US-08-975-166-8
Sequence 8, Application US/08975166
Patent No. 5928877
GENERAL INFORMATION:
APPLICANT: Tripp, Matthew
APPLICANT: Lusk, Lance
APPLICANT: Rhodes, Thomas
APPLICANT: Hulge, Nick
APPLICANT: Kot, Edward
APPLICANT: Chicoye, Etzer
APPLICANT: Barney, Michael C.
APPLICANT: Bower, Patricia A.
APPLICANT: Cronan, Charles L.
TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS
TITLE OF INVENTION: IN YEAST
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Thad Kryshak, Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MSWORD Version 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,166
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/180,524
FILING DATE:
APPLICATION NUMBER: US/07/917,216
FILING DATE:
APPLICATION NUMBER: US 07/486,333
FILING DATE: 28-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/409,217
FILING DATE: 19-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 66-005-9234-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5707
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-975-166-8

Query Match 4.1%; Score 7; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 13

US-08-975-166-9
; Sequence 9, Application US/08975166

Patent No. 5928877

GENERAL INFORMATION:

APPLICANT: Tripp, Matthew

APPLICANT: Lusk, Lance

APPLICANT: Rhodes, Thomas

APPLICANT: Hudge, Nick

APPLICANT: Kol, Edward

APPLICANT: Chlocoy, Elzer

APPLICANT: Barney, Michael C.

APPLICANT: Bower, Patricia A.

APPLICANT: Cronan, Charles L.

TITLE OF INVENTION: METHOD OF EXPRESSING ANTIFREEZE PROTEINS

TITLE OF INVENTION: IN YEAST

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thad Kryshak, Quarles & Brady

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: USA

ZIP: 53202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/975,166

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/180,524

FILING DATE:

APPLICATION NUMBER: US/07/917,216

FILING DATE:

APPLICATION NUMBER: US 07/486,333

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/409,217

FILING DATE: 19-SEP-1989

ATTORNEY/AGENT INFORMATION:

NAME: Kryshak, Thad

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 66-005-9234-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (414) 277-5707

TELEFAX: (414) 277-5591

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 37 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYDROPHETICAL: NO

ANTI-SENSE: NO

US-08-975-166-9

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 14

US-09-117-121-39
; Sequence 39, Application US/09117121

Patent No. 6307020

GENERAL INFORMATION:

APPLICANT: Hew, Choy

APPLICANT: Gong, Zhiyuan

TITLE OF INVENTION: Intracellular Antifreeze Polypeptides

TITLE OF INVENTION: and Nucleic Acids

NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/117,121

FILING DATE: 20-NOV-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/CA97/00062

FILING DATE: 30-JAN-1997

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 016252-00161005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:

LENGTH: 37 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: -

LOCATION: 1..37

OTHER INFORMATION: /note="HPLC-6"

US-09-117-121-39

Query Match 4.1%; Score 7; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 15

US-09-117-121-42
; Sequence 42, Application US/09117121

Patent No. 6307020

GENERAL INFORMATION:

APPLICANT: Hew, Choy

APPLICANT: Gong, Zhiyuan

TITLE OF INVENTION: Intracellular Antifreeze Polypeptides

TITLE OF INVENTION: and Nucleic Acids
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/117,121
FILING DATE: 20-NOV-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/CA97/00062
FILING DATE: 30-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 016252-00161005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-117-121-42

Query Match 4.1%; Score 7; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 6 AAAAAL 12

RESULT 16
US-09-344-529-4
Sequence 4, Application US/09344529
Patent No. 6429293
GENERAL INFORMATION:
APPLICANT: Hew, Choy L.
APPLICANT: HSC Research and Development Limited Partnership
TITLE OF INVENTION: Sculpin-Type Antifreeze Polypeptides and Nucleic Acids
FILE REFERENCE: 016252-002620US
CURRENT APPLICATION NUMBER: US/09/344,529
CURRENT FILING DATE: 1999-06-24
EARLIER APPLICATION NUMBER: US 60/090,794
EARLIER FILING DATE: 1998-06-26
EARLIER APPLICATION NUMBER: US 60/095,713
EARLIER FILING DATE: 1998-08-07
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 37
TYPE: PRT
ORGANISM: Pleuronectes americanus
FEATURE:
OTHER INFORMATION: Winter flounder liver-type antifreeze polypeptide
OTHER INFORMATION: (wLIAPF-6)
US-09-344-529-4

Query Match 4.1%; Score 7; DB 4; Length 37;
Best Local Similarity 100.0%; Pred. No. 6;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 6 AAAAAL 12

RESULT 17
US-07-814-220-1
Sequence 1, Application US/07814220
Patent No. 5925540
GENERAL INFORMATION:
APPLICANT: Caceci, Thomas
APPLICANT: Toth, Thomas E.
APPLICANT: Szumanski, Maria B.W.
TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND
TITLE OF INVENTION: SYNTHETIC GENE CODING FOR ITS PRODUCTION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: WHITAM, CURTIS & WHITAM
STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,
CITY: Reston
STATE: VA
COUNTRY: USA
ZIP: 20191
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/814,220
FILING DATE: 23-DEC-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/588,437
FILING DATE: 25-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Whitlam, Michael E.
REGISTRATION NUMBER: 32,635
REFERENCE/DOCKET NUMBER: CIT. 016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-391-2510
TELEFAX: 703-391-9035
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-814-220-1

Query Match 4.1%; Score 7; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 6 AAAAAL 12

RESULT 18
US-07-812-421-1
Sequence 1, Application US/07812421
Patent No. 5932697
GENERAL INFORMATION:
APPLICANT: Caceci, Thomas
APPLICANT: Toth, Thomas E.
APPLICANT: Szumanski, Maria B.W.
TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND
TITLE OF INVENTION: SYNTHETIC GENE CODING FOR ITS PRODUCTION

NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: WHITAM, CURTIS & WHITAM
STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,
CITY: Reston
STATE: VA
COUNTRY: USA
ZIP: 20191
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/812,421
FILING DATE: 23-DEC-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/588,437
FILING DATE: 25-SEP-1990
ATTORNEY/AGENT INFORMATION:
NAME: Whitlam, Michael E.
REGISTRATION NUMBER: 32,635
REFERENCE/DOCKET NUMBER: CIT.016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-391-2510
TELEFAX: 703-391-9035
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-812-421-1

Query Match
Best Local Similarity 4.1%; Score 7; DB 2; Length 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAL 14
Db 6 AAAAAAL 12

RESULT 19
US-09-325-932A-144
Sequence 144, Application US/09325932A
Patent No. 6451604
GENERAL INFORMATION:
APPLICANT: Filinn, Barry
APPLICANT: Lasham, Annette
TITLE OF INVENTION: death and their use in the modification of forestry plant develop
FILE REFERENCE: 1022
CURRENT APPLICATION NUMBER: US/09/325,932A
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 206
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 144
LENGTH: 109
TYPE: PRT
ORGANISM: Eucalyptus grandis
US-09-325-932A-144

Query Match
Best Local Similarity 4.1%; Score 7; DB 4; Length 109;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAAL 14
Db 18 AAAAAAL 24

RESULT 20
US-08-233-609-3
Sequence 3, Application US/08233609
Patent No. 5534615
GENERAL INFORMATION:
APPLICANT: Baker, Joffre
APPLICANT: Chien, Kenneth
APPLICANT: King, Kathleen
APPLICANT: Pennica, Diane
APPLICANT: Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,609
FILING DATE: 25-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 894
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/952-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-233-609-3

Query Match
Best Local Similarity 4.1%; Score 7; DB 1; Length 203;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALSYL 17
Db 91 AALSYL 97

RESULT 21
US-08-444-083-3
Sequence 3, Application US/08444083
Patent No. 5571675
GENERAL INFORMATION:
APPLICANT: Baker, Joffre
APPLICANT: Chien, Kenneth
APPLICANT: King, Kathleen
APPLICANT: Pennica, Diane
APPLICANT: Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco

STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,083
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894P1D5
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-444-083-3

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALSVL 17
Db 91 AAALSVL 97

RESULT 22
US-08-286-304-3
Sequence 3, Application US/08286304
Patent No. 5571893
GENERAL INFORMATION:
APPLICANT: Baker, Joffre
APPLICANT: Chien, Kenneth
APPLICANT: King, Kathleen
APPLICANT: Pennica, Diane
APPLICANT: Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,304
FILING DATE: 05-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994

ATTORNEY/AGENT INFORMATION:
NAME: Haasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 894P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-286-304-3

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AAALSVL 17
Db 91 AAALSVL 97

RESULT 23
US-08-442-745-3
Sequence 3, Application US/08442745
Patent No. 5624806
GENERAL INFORMATION:
APPLICANT: Baker, Joffre
APPLICANT: Chien, Kenneth
APPLICANT: King, Kathleen
APPLICANT: Pennica, Diane
APPLICANT: Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,745
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894P1D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-442-745-3

Query Match 4.1%; Score 7; DB 1;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALSVL 17
Db 91 AALSVL 97

RESULT 24

US-08-443-129-3
; Sequence 3, Application US/08443129
; Patent No. 5627073
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,129
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/233609
; FILING DATE: 25-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/286
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: 304
; FILING DATE: (null)
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 894P1D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-443-129-3

Query Match 4.1%; Score 7; DB 1;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALSVL 17
Db 91 AALSVL 97

RESULT 25
US-08-443-952-3
; Sequence 3, Application US/08443952

Patent No. 5679545
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,952
; FILING DATE: 17-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/233609
; FILING DATE: 25-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/286304
; FILING DATE: 05-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Torchia, Timothy E.
; REGISTRATION NUMBER: 36,700
; REFERENCE/DOCKET NUMBER: 894P1D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8674
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-443-952-3

Query Match 4.1%; Score 7; DB 1;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 AALSVL 17
Db 91 AALSVL 97

RESULT 26

US-08-443-130-3
; Sequence 3, Application US/08443130
; Patent No. 5723585
; GENERAL INFORMATION:
; APPLICANT: Baker, Joffre
; APPLICANT: Chien, Kenneth
; APPLICANT: King, Kathleen
; APPLICANT: Pennica, Diane
; APPLICANT: Wood, William
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California

COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,130
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: 894P1D3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8674
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-443-130-3

Query Match 4.1%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
| | | | |
Db 91 AALSVL 97

RESULT 27
US-08-898-911-3
Sequence 3, Application US/08898911
Patent No. 6117650
GENERAL INFORMATION:
APPLICANT: King, Kathleen
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/898,911
FILING DATE: 23-Jul-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/452555

FILING DATE: 25-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: P0894P1D6C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-2066
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-898-911-3

Query Match 4.1%; Score 7; DB 3; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
| | | | |
Db 91 AALSVL 97

RESULT 28
PCT-US95-04467-3
Sequence 3, Application PC/TUS9504467
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
TITLE OF INVENTION: Cardiotrophin and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04467
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/233609
FILING DATE: 25-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 894P1PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US95-04467-3

Query Match 4.1%; Score 7; DB 5; Length 203;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AALSVL 17
| | | | |
Db 91 AALSVL 97

RESULT 29
US-09-286-690-12
; Sequence 12, Application US/09286690
; Patent No. 6103511
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; TITLE OF INVENTION: Lichense and Coding Sequences
; FILE REFERENCE: 55-96
; CURRENT APPLICATION NUMBER: US/09/286,690
; EARLIER FILING DATE: 1999-04-05
; EARLIER APPLICATION NUMBER: US 60/027,882
; EARLIER FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: PCT/US97/17811
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 12
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Fibrobacter succinogenes
US-09-286-690-12

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 228;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 15 AAAAAA 21

RESULT 30
US-08-508-761B-32
; Sequence 32, Application US/08508761B
; Patent No. 6027920
; GENERAL INFORMATION:
; APPLICANT: Joliff, Gwennael
; APPLICANT: Guyonvarch, Armel
; APPLICANT: Purification, Relano
; APPLICANT: Duchiron, Francis
; APPLICANT: Renaud, Michel
; TITLE OF INVENTION: System for Protein Expression and
; TITLE OF INVENTION: Secretion Especially in Corynebacteria
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
; STREET: 400 Seventh St. N.W.
; CITY: Washington D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,761B
; FILING DATE: 31-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/09652
; FILING DATE: 29-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91/09870
; FILING DATE: 02-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Player, William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P58525NA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666

TELEFAX: (202) 393-5350
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: M. kansasii
US-08-508-761B-32

Query Match
Best Local Similarity 4.1%; Score 7; DB 3; Length 309;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 3 AAAAAA 9

RESULT 31
US-08-107-676-30
; Sequence 30, Application US/08107676
; Patent No. 5955356
; GENERAL INFORMATION:
; APPLICANT: Content, Jean
; APPLICANT: Demit, Lucas
; APPLICANT: Debruy, Jacqueline
; TITLE OF INVENTION: Mycobacterium polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding Them for Diagnosis and Control of
; TITLE OF INVENTION: Tuberculosis
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5955356west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,676
; FILING DATE: 04-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP92/00268
; FILING DATE: 07-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91400388.4
; FILING DATE: 14-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 8076.89USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium kansasii
; IMMEDIATE SOURCE:
; CLONE: Partial protein sequence from antigen 85B
; CLONE: from M.kansasii

US-08-107-676-30

Query Match 4.1%; Score 7; DB 2; Length 325;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
111111
DB 19 AAAAAA 25

RESULT 32

US-09-161-241-9
; Sequence 9, Application US/09161241
; Patent No. 6344541
; GENERAL INFORMATION:
; APPLICANT: Bass, Michael B
; APPLICANT: Sullivan, John K
; APPLICANT: Theill, Lars E
; APPLICANT: Wang, Daquan
; TITLE OF INVENTION: NOVEL DKR POLYPEPTIDES
; FILE REFERENCE: A-548
; CURRENT APPLICATION NUMBER: US/09/161,241
; CURRENT FILING DATE: 1998-09-25
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Human
US-09-161-241-9

Query Match 4.1%; Score 7; DB 4; Length 350;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
111111
DB 338 AAAAAA 344

RESULT 33

US-07-792-885A-1
; Sequence 1, Application US/07792885A
; Patent No. 551651
; GENERAL INFORMATION:
; APPLICANT: Goldring, Steven R.
; APPLICANT: Gorn, Alan H.
; APPLICANT: Lin, Herb Y.
; TITLE OF INVENTION: MAMMALIAN CALCITONIN RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,885A
; FILING DATE: 19911115
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00786/074001

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

LENGTH: 482
; TYPE: AMINO ACID
; STRANDEDNESS: N/A
; TOPOLOGY: N/A
US-07-792-885A-1

Query Match 4.1%; Score 7; DB 1; Length 482;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
111111
DB 431 AAAAAA 437

RESULT 34

US-08-142-439A-7
; Sequence 7, Application US/08142439A
; Patent No. 5670360
; GENERAL INFORMATION:
; APPLICANT: Thorens, Bernard
; TITLE OF INVENTION: Receptor for the Glucagon-Like-Peptide-1
; TITLE OF INVENTION: (GLP-1)
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 56703600 No. 5670360disk of No. 5670360th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/142,439A

FILING DATE: 24-NOV-93

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 398/92

FILING DATE: 25-MAR-92

PRIOR APPLICATION DATA: PCT/EP93/00697

APPLICATION NUMBER: PCT/EP93/00697

FILING DATE: 23-MAR-93

ATTORNEY/AGENT INFORMATION:

NAME: Harrington, James J.

REGISTRATION NUMBER: 38,711

REFERENCE/DOCKET NUMBER: 3756, 204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 482 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Sus scrofa

US-08-142-439A-7

Query Match 4.18; Score 7; DB 1; Length 482;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 431 AAAAAL 437

RESULT 35
US-08-869-477-7
Sequence 7, Application US/08869477
Patent No. 5846747
GENERAL INFORMATION:
APPLICANT: Thorens, Bernard
TITLE OF INVENTION: Receptor for the Glucagon-Like-Peptide-1
TITLE OF INVENTION: (GLP-1)
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5846747, No. 5846747, disk of No. 5846747, America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/869,477
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/142,439
FILING DATE: 24-NOV-93
APPLICATION NUMBER: DK 398/92
FILING DATE: 25-MAR-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/00697
FILING DATE: 23-MAR-93
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 3756,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 482 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HIPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
US-08-869-477-7

Query Match 4.18; Score 7; DB 2; Length 482;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
Db 431 AAAAAL 437

RESULT 36
US-09-172-952-26

Sequence 26, Application US/09172952
Patent No. 6368793
GENERAL INFORMATION:
APPLICANT: Hoch, James
APPLICANT: Dattois, Veronique
TITLE OF INVENTION: METABOLIC SELECTION METHODS
FILE REFERENCE: 234/191
CURRENT APPLICATION NUMBER: US/09/172,952
CURRENT FILING DATE: 1998-10-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 26
LENGTH: 498
TYPE: PRT
ORGANISM: Lyxk-Ec
US-09-172-952-26

Query Match 4.18; Score 7; DB 4; Length 498;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 YNMSLGE 122
Db 184 YNMSLGE 190

RESULT 37
US-09-382-256-10
Sequence 10, Application US/09382256A
Patent No. 6207814
GENERAL INFORMATION:
APPLICANT: MIYAZONO, Kohel
TEN DIKE, Peter
FRANZEN, Petra
YAMASHITA, Hidetoshi
HELDIN, Carl-Henrik
TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS
HAVING SERINE THREONINE KINASE DOMAINS,
AND THEIR USE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/382,256A
FILING DATE: 24-Aug-1999
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: No. 6207814, December 17, 1993
APPLICATION NUMBER: GB 9224057.1
FILING DATE: No. 6207814, December 17, 1992
APPLICATION NUMBER: GB 9304677.9
FILING DATE: March 8, 1993
APPLICATION NUMBER: GB 9304680.3
FILING DATE: March 8, 1993
APPLICATION NUMBER: 9311047.6
FILING DATE: May 28, 1993
APPLICATION NUMBER: 9313763.6
FILING DATE: July 2, 1993
APPLICATION NUMBER: 9316099.2
FILING DATE: August 3, 1993
APPLICATION NUMBER: 321344.5
FILING DATE: October 15, 1993

ATTORNEY/AGENT INFORMATION:
NAME: NO. 6207814man D. Hanson
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5298.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-382-256-10

Query Match 4.1%; Score 7; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
| | | | |
Db 21 AAAAAA 27

RESULT 38
US-09-395-115-10
Sequence 10, Application US/09395115
Patent No. 6271365
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/395,115
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6271365ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6271365ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohel, Vlnet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-395-115-10

Query Match 4.1%; Score 7; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
| | | | |
Db 21 AAAAAA 27

RESULT 39
US-08-436-265-10
Sequence 10, Application US/08436265
Patent No. 6316217
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felle & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6316217ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6316217ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
TELECOMMUNICATION INFORMATION: LUD 5298
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-265-10

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 503;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
DB 21 AAAAAL 27

RESULT 40
US-09-679-187-10
Sequence 10, Application US/09679187
Patent No. 6331621
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohlei, Peter Ten;
APPLICANT: Franzen, Petrus, Yamashita, Hidekoshi; Heidin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESS: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/679,187
FILING DATE: 03-OCT-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-NO. 6331621ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-NO. 6331621ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:

```

```

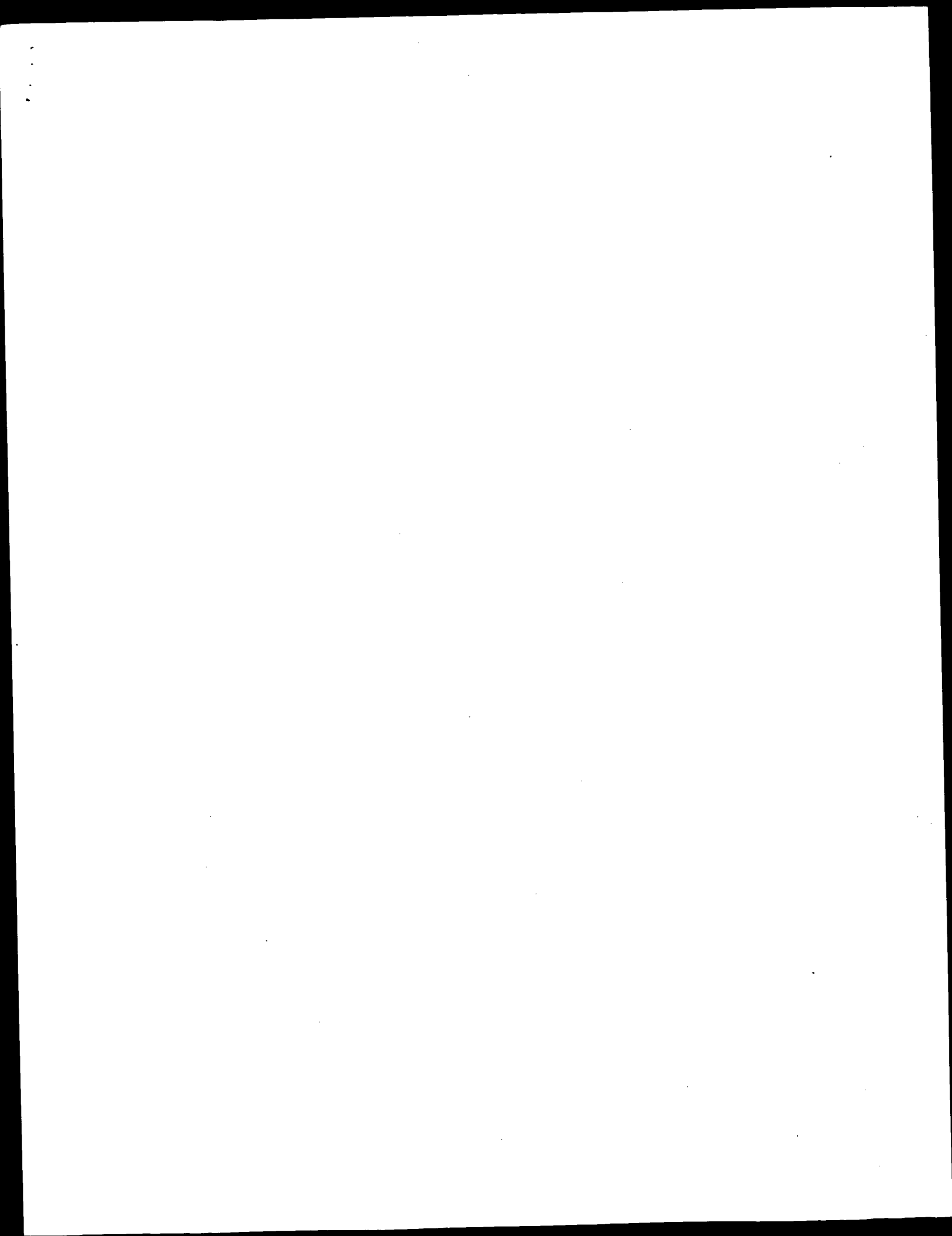
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
TELECOMMUNICATION INFORMATION: LUD 5298
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-679-187-10

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 503;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAL 14
DB 21 AAAAAL 27

Search completed: July 6, 2003, 14:22:34
Job time : 21 secs

```



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:22:40 ; Search time 24 Seconds

(without alignments)
824.116 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 172
Sequence: 1 MMLHIQIAAAALSVLFM.....IAFGTNEAMSONRAELSY 172

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 440863 seqs, 114992915 residues

Word size : 0

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 100 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	4.7	1744	9 US-10-108-605-25	Sequence 25, Appl
2	7	4.1	21	10 US-09-864-761-42679	Sequence 42679, A
3	7	4.1	22	9 US-10-024-450-7	Sequence 7, Appl
4	7	4.1	23	9 US-10-311-111-29	Sequence 165, App
5	7	4.1	38	9 US-09-776-724A-165	Sequence 1, Appl
6	7	4.1	41	9 US-10-106-487-1	Sequence 40439, A
7	7	4.1	52	10 US-09-864-761-40439	Sequence 10, Appl
8	7	4.1	78	10 US-09-795-501-10	Sequence 770, App
9	7	4.1	79	9 US-10-092-154-770	Sequence 144, App
10	7	4.1	79	9 US-09-764-847-770	Sequence 2737, Ap
11	7	4.1	109	9 US-10-219-220-144	Sequence 44, Appl
12	7	4.1	119	9 US-09-764-891-2737	Sequence 49, Appl
13	7	4.1	131	10 US-09-164-615-44	Sequence 42364, A
14	7	4.1	132	10 US-09-164-615-49	Sequence 2, Appl
15	7	4.1	161	10 US-09-864-761-42364	Sequence 3, Appl
16	7	4.1	192	9 US-09-747-348-2	Sequence 36691, A
17	7	4.1	203	9 US-10-107-931-3	
18	7	4.1	203	10 US-09-896-856-3	
19	7	4.1	273	10 US-09-864-761-36691	

20	7	4.1	287	9 US-10-156-761-9684	Sequence 9684, Ap
21	7	4.1	300	9 US-10-156-761-12279	Sequence 12279, A
22	7	4.1	319	9 US-10-156-761-9015	Sequence 9015, Ap
23	7	4.1	324	9 US-10-281-024-15	Sequence 15, Appl
24	7	4.1	340	9 US-10-156-761-9858	Sequence 9858, Ap
25	7	4.1	350	9 US-09-905-291A-236	Sequence 236, Appl
26	7	4.1	350	9 US-09-976-736-9	Sequence 9, Appl
27	7	4.1	350	9 US-10-063-547-8	Sequence 8, Appl
28	7	4.1	350	9 US-09-902-853-236	Sequence 236, App
29	7	4.1	350	9 US-09-907-824-236	Sequence 236, App
30	7	4.1	350	9 US-09-907-841-236	Sequence 236, App
31	7	4.1	350	9 US-09-904-011-236	Sequence 236, App
32	7	4.1	350	9 US-10-201-310-2	Sequence 2, Appl
33	7	4.1	350	9 US-10-063-616-8	Sequence 8, Appl
34	7	4.1	350	9 US-09-906-742-236	Sequence 236, Appl
35	7	4.1	350	9 US-10-063-502-8	Sequence 8, Appl
36	7	4.1	350	9 US-09-906-838-236	Sequence 236, App
37	7	4.1	350	9 US-09-907-613-236	Sequence 236, App
38	7	4.1	350	9 US-09-907-842-236	Sequence 236, App
39	7	4.1	350	9 US-09-904-820-236	Sequence 236, App
40	7	4.1	350	9 US-09-904-859-236	Sequence 236, App
41	7	4.1	350	9 US-09-909-204-236	Sequence 236, App
42	7	4.1	350	9 US-09-904-786-236	Sequence 236, App
43	7	4.1	350	9 US-09-906-646-236	Sequence 236, App
44	7	4.1	350	9 US-09-906-700-236	Sequence 236, App
45	7	4.1	350	9 US-09-902-903-236	Sequence 236, App
46	7	4.1	350	9 US-09-903-749A-236	Sequence 236, App
47	7	4.1	350	9 US-09-903-786-236	Sequence 236, App
48	7	4.1	350	9 US-09-902-736-236	Sequence 236, App
49	7	4.1	350	9 US-09-904-119-236	Sequence 236, App
50	7	4.1	350	9 US-09-904-956-236	Sequence 236, App
51	7	4.1	350	9 US-09-907-794-236	Sequence 236, App
52	7	4.1	350	9 US-10-063-518-8	Sequence 8, Appl
53	7	4.1	350	9 US-10-063-598-8	Sequence 8, Appl
54	7	4.1	350	9 US-10-227-693-8	Sequence 8, Appl
55	7	4.1	350	9 US-09-902-692-236	Sequence 236, App
56	7	4.1	350	9 US-09-903-520-236	Sequence 236, App
57	7	4.1	350	9 US-09-903-943-236	Sequence 236, App
58	7	4.1	350	9 US-09-904-462-236	Sequence 236, App
59	7	4.1	350	9 US-09-905-056-236	Sequence 236, App
60	7	4.1	350	9 US-09-907-925-236	Sequence 236, App
61	7	4.1	350	9 US-09-904-553-236	Sequence 236, App
62	7	4.1	350	9 US-09-905-381-236	Sequence 236, App
63	7	4.1	350	9 US-09-909-064-236	Sequence 236, App
64	7	4.1	350	9 US-09-972-473-2	Sequence 2, Appl
65	7	4.1	350	9 US-10-063-567-8	Sequence 8, Appl
66	7	4.1	350	9 US-09-905-088-236	Sequence 236, App
67	7	4.1	350	9 US-09-907-575-236	Sequence 236, App
68	7	4.1	350	9 US-10-063-538-8	Sequence 8, Appl
69	7	4.1	350	9 US-09-902-759-236	Sequence 236, App
70	7	4.1	350	9 US-09-905-075-236	Sequence 236, App
71	7	4.1	350	9 US-10-063-599-8	Sequence 8, Appl
72	7	4.1	350	9 US-09-902-634-236	Sequence 236, App
73	7	4.1	350	9 US-09-902-713-236	Sequence 236, App
74	7	4.1	350	9 US-09-907-979-236	Sequence 236, App
75	7	4.1	350	9 US-10-063-595-8	Sequence 8, Appl
76	7	4.1	350	9 US-09-902-615-236	Sequence 236, App
77	7	4.1	350	9 US-09-903-925-236	Sequence 236, App
78	7	4.1	350	9 US-09-906-760A-236	Sequence 236, App
79	7	4.1	350	9 US-10-223-088-50	Sequence 50, Appl
80	7	4.1	350	9 US-10-271-628-3	Sequence 3, Appl
81	7	4.1	350	9 US-09-903-823-236	Sequence 236, App
82	7	4.1	350	9 US-09-907-652-236	Sequence 236, App
83	7	4.1	350	9 US-10-063-580-8	Sequence 8, Appl
84	7	4.1	350	9 US-10-223-084-50	Sequence 50, Appl
85	7	4.1	350	9 US-10-223-088-50	Sequence 50, Appl
86	7	4.1	350	9 US-10-223-090-50	Sequence 50, Appl
87	7	4.1	350	9 US-09-902-572A-236	Sequence 236, App
88	7	4.1	350	9 US-10-063-557-8	Sequence 8, Appl
89	7	4.1	350	9 US-10-223-087-50	Sequence 50, Appl
90	7	4.1	350	9 US-09-902-979-236	Sequence 236, App
91	7	4.1	350	9 US-09-905-125-236	Sequence 236, App
92	7	4.1	350	9 US-09-906-815A-236	Sequence 236, App

93 7 4.1 350 9 US-10-063-585-8 Sequence 8, Appl
94 7 4.1 350 9 US-10-223-083-50 Sequence 50, Appl
95 7 4.1 350 10 US-09-909-320-236 Sequence 236, App
96 7 4.1 350 10 US-09-909-088B-236 Sequence 236, App
97 7 4.1 350 12 US-10-006-867-8 Sequence 8, Appl
98 7 4.1 356 10 US-09-815-242-4875 Sequence 4875, Ap
99 7 4.1 356 10 US-09-815-242-10782 Sequence 10782, A
100 7 4.1 395 9 US-10-153-668-330 Sequence 330, App

ALIGNMENTS

RESULT 1
US-10-108-605-25
Sequence 25, Application US/10108605
Patent No. US20020160934A1
GENERAL INFORMATION:
APPLICANT: Brodus, Julie
APPLICANT: Bachmann, Lynn
APPLICANT: Kamdar, Kim
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
FILE REFERENCE: 31133B
CURRENT APPLICATION NUMBER: US/10/108, 605
PRIOR FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: US 09/761,142
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/176,418
PRIOR FILING DATE: 2000-01-14
NUMBER OF SEQ ID NOS: 361
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 25
LENGTH: 1744
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-10-108-605-25
Query Match 4.7%; Score 8; DB 9; Length 1744;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 AAAAAALS 15
DB 941 AAAAAALS 948
RESULT 2
US-09-864-761-42679
Sequence 42679, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263, 6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42679
LENGTH: 21
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007567.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.86
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3
OTHER INFORMATION: EST_HUMAN HIT: AL134791.1, EVALUATE 1.00e-03
US-09-864-761-42679
Query Match 4.1%; Score 7; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 130 NYLKGK 136
DB 4 NYLKGK 10
RESULT 3
US-10-024-450-7
Sequence 7, Application US/10024450
Publication No. US20030032606A1
GENERAL INFORMATION:
APPLICANT: Huang, Shi
APPLICANT: Chadwick, Robert B.
TITLE OF INVENTION: Methods of Detecting and Treating
FILE REFERENCE: P-1J 5101
CURRENT APPLICATION NUMBER: US/10/024,450
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: US 60/256,582
PRIOR FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 22
TYPE: PRT
ORGANISM: Homo sapiens

US-10-024-450-7

Query Match

Best Local Similarity 4.1%; Score 7; DB 9; Length 22;
Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 8 AAAAAA 14
|||||

DB 6 AAAAAA 12

RESULT 4

US-10-311-111-29

Sequence 29, Application US/10311111
Publication No. US20030121065A1

GENERAL INFORMATION:

APPLICANT: SHIBA, KIYOTAKA

TITLE OF INVENTION: MULTIFUNCTIONAL BASE SEQUENCE AND ARTIFICIAL GENE CONTAINING THE

FILE REFERENCE: 4439-4004

CURRENT FILING DATE: 2002-12-13

PRIOR APPLICATION NUMBER: JP 2000-180997

PRIOR FILING DATE: 2000-06-16

NUMBER OF SEQ ID NOS: 34

SOFTWARE: PatentIn version 3.1

SEQ ID NO 29

LENGTH: 23

TYPE: PRT

ORGANISM: artificial

FEATURE:

OTHER INFORMATION: Designed peptide

US-10-311-111-29

Query Match

Best Local Similarity 4.1%; Score 7; DB 9; Length 23;
Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 8 AAAAAA 14
|||||

DB 6 AAAAAA 12

RESULT 5

US-09-776-724A-165

Sequence 165, Application US/09776724A
Publication No. US20030050455A1

GENERAL INFORMATION:

APPLICANT: ROSEN ET AL.

TITLE OF INVENTION: 64 Human Secreted Proteins

FILE REFERENCE: P2011

CURRENT APPLICATION NUMBER: US/09/776,724A

CURRENT FILING DATE: 2001-02-06

PRIOR APPLICATION NUMBER: 60/180,909

PRIOR FILING DATE: 2000-02-08

PRIOR APPLICATION NUMBER: 09/669,688

PRIOR FILING DATE: 2000-09-26

PRIOR APPLICATION NUMBER: 09/229,982

PRIOR FILING DATE: 1999-01-14

PRIOR APPLICATION NUMBER: PCT/US98/14613

PRIOR FILING DATE: 1998-07-15

PRIOR APPLICATION NUMBER: 60/052,661

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,872

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,871

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,874

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,873

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,870

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/052,875

PRIOR FILING DATE: 1997-07-16

PRIOR APPLICATION NUMBER: 60/053,440

PRIOR FILING DATE: 1997-07-22

PRIOR APPLICATION NUMBER: 60/053,441

PRIOR FILING DATE: 1997-07-22

PRIOR APPLICATION NUMBER: 60/053,442

PRIOR FILING DATE: 1997-07-22

PRIOR APPLICATION NUMBER: 60/056,359

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,725

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,985

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,952

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,989

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/056,361

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,726

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,724

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,946

PRIOR FILING DATE: 1997-08-18

PRIOR APPLICATION NUMBER: 60/055,683

PRIOR FILING DATE: 1997-08-18

NUMBER OF SEQ ID NOS: 294

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 165

LENGTH: 38

TYPE: PRT

ORGANISM: Homo sapiens

US-09-776-724A-165

Query Match

Best Local Similarity 4.1%; Score 7; DB 9; Length 38;
Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 8 AAAAAA 14
|||||

DB 23 AAAAAA 29

RESULT 6

US-10-106-487-1

Sequence 1, Application US/10106487
Patent No. US20020164721A1

GENERAL INFORMATION:

APPLICANT: FIRAT, HOSEYIN

APPLICANT: LEMONNIER, FRANCOIS

APPLICANT: LANGLADE-DEMOYEN, PIERRE

APPLICANT: MICHEL, MARIE-LOUISE

TITLE OF INVENTION: DESIGN OF A POLYPEPTIDIC CONSTRUCT FOR THE INDUCTION

TITLE OF INVENTION: OF

TITLE OF INVENTION: HLA-A2.1 RESTRICTED HIV 1 SPECIFIC CTL RESPONSES USING

FILE REFERENCE: 03495.0196 SEQUENCE LISTING

CURRENT APPLICATION NUMBER: US/10/106,487

CURRENT FILING DATE: 2002-03-27

PRIOR APPLICATION NUMBER: 09/675,673

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/158,356

PRIOR FILING DATE: 1999-10-12

NUMBER OF SEQ ID NOS: 41

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 41

TYPE: PRT

ORGANISM: Hepatitis B virus

US-10-106-487-1

Query Match

4.1%; Score 7; DB 9; Length 41;

Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 6 AAAAAA 12

RESULT 7

US-09-864-761-40439
; Sequence 40439, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40439
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121747.19
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EST_HUMAN HIT: A0131227.1, EVALUATE 3.00e-05
US-09-864-761-40439

Query Match 4.1%; Score 7; DB 10; Length 52;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 7 AAAAAA 13

RESULT 8

US-09-795-501-10
; Sequence 10, Application US/09795501
; Patent No. US20020042098A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32253 TRANSFERASE FAMILY MEMBERS AND
; FILE REFERENCE: 38155-20004.00
; CURRENT FILING DATE: 2001-02-28
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-09-795-501-10

Query Match 4.1%; Score 7; DB 10; Length 78;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 60 AAAAAA 66

RESULT 9

US-10-092-154-770
; Sequence 770, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT FILING DATE: 2002-03-07
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See file wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 770
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-092-154-770

Query Match 4.1%; Score 7; DB 9; Length 79;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAA 13
Db 42 IAAAAA 48

RESULT 10
US-09-764-847-770
; Sequence 770, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 770
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-847-770

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAA 13
Db 42 IAAAAA 48

RESULT 11
US-10-219-220-144
; Sequence 144, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 144
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-10-219-220-144

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 109;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 18 AAAAAA 24

RESULT 12
US-09-764-891-2737
; Sequence 2737, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2737
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (103)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-2737

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 119;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 77 AAAAAA 83

RESULT 13
US-09-164-615-44
; Sequence 44, Application US/09164615B
; Patent No. US20020073442A1
; GENERAL INFORMATION:
; APPLICANT: Hanley-Bowdoin, Linda
; APPLICANT: Settlage, Sharon
; TITLE OF INVENTION: Geminiivirus Resistant Transgenic Plants
; FILE REFERENCE: 5051-433
; CURRENT APPLICATION NUMBER: US/09/164,615B
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: TYCLV mutant
US-09-164-615-44

Query Match
Best Local Similarity 4.1%; Score 7; DB 10; Length 131;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 IAAAAA 13
Db 48 IAAAAA 54

RESULT 14
US-09-164-615-49
; Sequence 49, Application US/09164615B
; Patent No. US20020073442A1
; GENERAL INFORMATION:
; APPLICANT: Hanley-Bowdoin, Linda
; APPLICANT: Settlage, Sharon
; TITLE OF INVENTION: Geminiivirus Resistant Transgenic Plants
; FILE REFERENCE: 5051-433
; CURRENT APPLICATION NUMBER: US/09/164,615B
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 132

;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: TGMV AL3
;; OTHER INFORMATION: mutant (MAL3#67)
US-09-164-615-49

Query Match 4.1%; Score 7; DB 10; Length 132;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 49 AAAAAA 55

RESULT 15
US-09-864-761-42364
Sequence 42364, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42364
LENGTH: 161
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

;; OTHER INFORMATION: MAP TO AC011447.2
;; OTHER INFORMATION: EXPRESSED IN PLENTA, SIGNAL = 3.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.8
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.9
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
;; OTHER INFORMATION: SWISSPROT HIT: P11227, EVALUATE 9.00e-14
;; OTHER INFORMATION: EST_HUMAN HIT: AW753028.1, EVALUATE 7.00e-61
US-09-864-761-42364

Query Match 4.1%; Score 7; DB 10; Length 161;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
DB 7 AAAAAA 13

RESULT 16
US-09-747-348-2
Sequence 2, Application US/09747348
Patent No. US20020123067A1

GENERAL INFORMATION:
APPLICANT: Andrew D. MURDIN, Raymond P. OOMEN, Joe WANG, Pamela DUNN
TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses th
FILE REFERENCE: 77813-37
CURRENT APPLICATION NUMBER: US/09/747,348
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 60/171,525
PRIOR FILING DATE: 1999-12-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 192
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-747-348-2

Query Match 4.1%; Score 7; DB 10; Length 192;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 GHTDERG 112
DB 123 GHTDERG 129

RESULT 17
US-10-107-931-3
Sequence 3, Application US/10107931
Publication No. US20030054550A1
GENERAL INFORMATION:
APPLICANT: Baker, Joffie
Chien, Kenneth
King, Kathleen
Pennica, Diane
Wood, William

TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/107,931
FILING DATE: 26-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/364,187
FILING DATE: 30-Jul-1999
APPLICATION NUMBER: 08/286,304
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 894P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-107-931-3

Query Match 4.1%; Score 7; DB 9; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALSVL 17
Db 91 AAALSVL 97

RESULT 18
US-09-896-856-3
Sequence 3, Application US/09896856
Patent No. US20020137189A1
GENERAL INFORMATION:
APPLICANT: Baker, Joffe
Chien, Kenneth
King, Kathleen
Pennica, Diane
Wood, William
TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/896,856
FILING DATE: 29-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/733,850
FILING DATE: 18-OCT-1996
APPLICATION NUMBER: US 08/471,112
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: 08/233,609
FILING DATE: 25-APR-1994
APPLICATION NUMBER: 08/286304
FILING DATE: 05-AUG-1994

APPLICATION NUMBER: 08/443129
FILING DATE: 17-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Conley, Deirdre L.
REGISTRATION NUMBER: 36,487
REFERENCE/DOCKET NUMBER: F0894PID2C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-2066
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-896-856-3

Query Match 4.1%; Score 7; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 AAALSVL 17
Db 91 AAALSVL 97

RESULT 19
US-09-864-761-36691
Sequence 36691, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
Applicant: Rank, David R.
Applicant: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21

```

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36691
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006460.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BT4/4, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: SWISSPROT HIT: 062722, EVALU0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: A0131373.1, EVALU0.100e-114
; US-09-864-761-36691

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 273;
Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 AAAALSV 16
    |||||
Db 193 AAAALSV 199

```

```

RESULT 20
US-10-156-761-9684
; Sequence 9684, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9684
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-9684

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 287;
Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 8 AAAAAL 14
    |||||
Db 157 AAAAAL 163

```

```

RESULT 21
US-10-156-761-12279

```

```

; Sequence 12279, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12279
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-12279

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 300;
Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 8 AAAAAL 14
    |||||
Db 43 AAAAAL 49

```

```

RESULT 22
US-10-156-761-9015
; Sequence 9015, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9015
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-9015

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 319;
Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 9 AAAAALS 15
    |||||
Db 301 AAAAALS 307

```

```

RESULT 23
US-10-281-024-15
; Sequence 15, Application US/10281024

```


Publication No. US20030087410A1
 GENERAL INFORMATION:
 APPLICANT: SATOSHI MORI
 APPLICANT: KYOKO HIGUCHI
 TITLE OF INVENTION: NICOTINAMINE SYNTHASE AND GENE ENCODING
 TITLE OF INVENTION: THE SAME
 FILE REFERENCE: 55107 (71526)
 CURRENT APPLICATION NUMBER: US/10/281,024
 CURRENT FILING DATE: 2002-10-25
 PRIOR APPLICATION NUMBER: US/09/674,337A
 PRIOR FILING DATE: 2002-01-07
 PRIOR APPLICATION NUMBER: PCT/JP99/02305
 PRIOR FILING DATE: 1999-04-30
 NUMBER OF SEQ ID NOS: 22
 SOFTWARE: FASTSEQ for Windows Version 3.0
 SEQ ID NO 15
 LENGTH: 324
 TYPE: PRT
 ORGANISM: Oryza sativa L.
 US-10-281-024-15

Query Match
 Best Local Similarity 4.1%; Score 7; DB 9; Length 324;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 63 ASKLPSTL 69
 Db 20 ASKLPSTL 26

RESULT 24
 US-10-156-761-9858
 Sequence 9858, Application US/10156761
 Publication No. US20030119018A1
 GENERAL INFORMATION:
 APPLICANT: OKURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 9858
 LENGTH: 340
 TYPE: PRT
 ORGANISM: Streptomyces avermitilis
 US-10-156-761-9858

Query Match
 Best Local Similarity 4.1%; Score 7; DB 9; Length 340;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 122 ERRVAV 128
 Db 199 ERRVAV 205

RESULT 25
 US-09-905-291A-236
 Sequence 236, Application US/09905291A
 Patent No. US20020160374A1
 GENERAL INFORMATION:
 APPLICANT: Genentech, Inc.
 APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David
 APPLICANT: Desnovers, Luc
 APPLICANT: Eaton, Dan L.
 APPLICANT: Ferrara, Napoleone
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerber, Hanspeter
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, A.
 APPLICANT: Grimaldi, Christopher J.
 APPLICANT: Gunney, Austin L.
 APPLICANT: Hillan, Kenneth, J.
 APPLICANT: Kljavin, Ivar J.
 APPLICANT: Mather, Jennie P.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Williams, P. Mickey
 APPLICANT: Wood, William, I.
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE REFERENCE: 10466-14
 CURRENT APPLICATION NUMBER: US/09/905,291A
 CURRENT FILING DATE: 2001-07-12
 PRIOR APPLICATION NUMBER: PCT/US00/04414
 PRIOR FILING DATE: 2000-02-22
 PRIOR APPLICATION NUMBER: US 60/143,048
 PRIOR FILING DATE: 1999-07-07
 PRIOR APPLICATION NUMBER: US 60/145,698
 PRIOR FILING DATE: 1999-07-26
 PRIOR APPLICATION NUMBER: US 60/146,222
 PRIOR FILING DATE: 1999-07-28
 PRIOR APPLICATION NUMBER: PCT/US99/20594
 PRIOR FILING DATE: 1999-09-08
 PRIOR APPLICATION NUMBER: PCT/US99/20944
 PRIOR FILING DATE: 1999-09-13
 PRIOR APPLICATION NUMBER: PCT/US99/21090
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/21547
 PRIOR FILING DATE: 1999-09-15
 PRIOR APPLICATION NUMBER: PCT/US99/23089
 PRIOR FILING DATE: 1999-10-05
 PRIOR APPLICATION NUMBER: PCT/US99/28214
 PRIOR FILING DATE: 1999-11-29
 PRIOR APPLICATION NUMBER: PCT/US99/28313
 PRIOR FILING DATE: 1999-11-30
 PRIOR APPLICATION NUMBER: PCT/US99/28564
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/28565
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: PCT/US99/30095
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: PCT/US99/30911
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US99/30999
 PRIOR FILING DATE: 1999-12-20
 PRIOR APPLICATION NUMBER: PCT/US00/00219
 PRIOR FILING DATE: 2000-01-05
 NUMBER OF SEQ ID NOS: 423
 SEQ ID NO 236
 LENGTH: 350
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-905-291A-236

Query Match
 Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 |||||
 Db 338 AAAAAA 344

RESULT 26
 US-09-976-736-9
 ; Sequence 9, Application US/09976736
 ; Patent No. US20020161178A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bass, Michael B
 ; APPLICANT: Sullivan, John K
 ; APPLICANT: Theill, Lars E
 ; APPLICANT: Wang, Daquan
 ; TITLE OF INVENTION: NOVEL DKR POLYPEPTIDES
 ; FILE REFERENCE: A-548
 ; CURRENT APPLICATION NUMBER: US/09/976,736
 ; CURRENT FILING DATE: 2001-10-09
 ; PRIOR APPLICATION NUMBER: US/09/161,241
 ; PRIOR FILING DATE: 1998-09-25
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 350
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-976-736-9

Query Match 4.1%; Score 7; DB 9; Length 350;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 |||||
 Db 338 AAAAAA 344

RESULT 27
 US-10-063-547-8
 ; Sequence 8, Application US/10063547
 ; Publication No. US20020182638A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gutney, Austin L.
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; TITLE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3230R1C1
 ; CURRENT APPLICATION NUMBER: US/10/063,547
 ; CURRENT FILING DATE: 2002-05-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; NUMBER OF SEQ ID NOS: 170
 ; SEQ ID NO 8
 ; LENGTH: 350
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-063-547-8

Query Match 4.1%; Score 7; DB 9; Length 350;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAA 14
 |||||
 Db 338 AAAAAA 344

RESULT 28
 US-09-902-853-236
 ; Sequence 236, Application US/09902853
 ; Publication No. US20020192659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Genentech, Inc.
 ; APPLICANT: Ashkenazi, Avi
 ; APPLICANT: Botstein, David
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Eaton, Dan L.
 ; APPLICANT: Ferrara, Napoleone
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Fong, Sherman
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerber, Hanspeter
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, A.
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, Christopher J.
 ; APPLICANT: Gutney, Austin L.
 ; APPLICANT: Hillan, Kenneth, J.
 ; APPLICANT: Kljavin, Ivar J.
 ; APPLICANT: Mather, Jennie P.
 ; APPLICANT: Pan, James
 ; APPLICANT: Paoni, Nicholas F.
 ; APPLICANT: Roy, Margaret Ann
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Williams, P. Mickey
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 ; TITLE OF INVENTION: Acids Encoding the Same
 ; FILE REFERENCE: 10466-14
 ; CURRENT APPLICATION NUMBER: US/09/902,853
 ; CURRENT FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: US/09/665,350
 ; PRIOR FILING DATE: 2000-09-18
 ; PRIOR APPLICATION NUMBER: US 60/143,048
 ; PRIOR FILING DATE: 1999-07-07
 ; PRIOR APPLICATION NUMBER: US 60/145,698
 ; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: US 60/146,222
 ; PRIOR FILING DATE: 1999-07-28
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-08
 ; PRIOR APPLICATION NUMBER: PCT/US99/20594
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/21090
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: PCT/US99/21547
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/23089
 ; PRIOR FILING DATE: 1999-10-05
 ; PRIOR APPLICATION NUMBER: PCT/US99/28214
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: PCT/US99/28313
 ; PRIOR FILING DATE: 1999-11-30
 ; PRIOR APPLICATION NUMBER: PCT/US99/28564
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/28565
 ; PRIOR FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: PCT/US99/30095
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: PCT/US99/30911
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US99/30999
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: PCT/US00/00219
 ; PRIOR FILING DATE: 2000-01-05
 ; NUMBER OF SEQ ID NOS: 423
 ; SEQ ID NO 236
 ; LENGTH: 350
 ; TYPE: PRT

ORGANISM: Homo Sapien
US-09-902-853-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
DB 338 AAAAAA 344

RESULT 29

US-09-907-824-236
Sequence 236, Application US/09907824
Publication No. US20020197671A1
GENERAL INFORMATION:

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertschen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gunney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,824
PRIOR FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Sapien
US-09-907-824-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
DB 338 AAAAAA 344

RESULT 30

US-09-907-841-236
Sequence 236, Application US/09907841
Publication No. US20020198366A1
GENERAL INFORMATION:

APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertschen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gunney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907,841
PRIOR FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13

```
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 236
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-841-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Pred. No. 1.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      8 AAAAAA 14
Db      338 AAAAAA 344

RESULT 31
US-09-904-011-236
; Sequence 236, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kiljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
```

```
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 236
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-904-011-236

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Pred. No. 1.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      8 AAAAAA 14
Db      338 AAAAAA 344

RESULT 32
US-10-201-310-2
; Sequence 2, Application US/10201310
; Publication No. US20030004327A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Ruden, Steven M.
; TITLE OF INVENTION: Cerebellum and Embryo Specific Protein
; FILE REFERENCE: 1488-0610002
; CURRENT APPLICATION NUMBER: US/10/201,310
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US 60/033,870
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: 08/993,198
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-201-310-2

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 350;
Pred. No. 1.9e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY      8 AAAAAA 14
Db      338 AAAAAA 344

RESULT 33
US-10-063-616-8
```

```
Sequence 8, Application US/10063616
; Publication No. US20030013855A1
; GENERAL INFORMATION:
; APPLICANT: Baton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,616
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 8
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-616-8

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      8 AAAAAA 14
Db      338 AAAAAA 344

RESULT 34
US-09-906-742-236
; Sequence 236, Application US/09906742
; Publication No. US20030023054A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kiljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paonli, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,742
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
```

```

; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 236
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-742-236

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      8 AAAAAA 14
Db      338 AAAAAA 344

RESULT 35
US-10-063-502-8
; Sequence 8, Application US/10063502
; Publication No. US20030023042A1
; GENERAL INFORMATION:
; APPLICANT: Baton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,502
; PRIOR FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 8
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapien
```

US-10-063-502-8

Query Match 4.1% Score 7: DB 9: Length 350;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
| | | | |
DB 338 AAAAAL 344

RESULT 36
US-09-906-838-236
; Sequence 236, Application US/09906838
; Publication No. US20030027143A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mathier, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,838
; PRIOR APPLICATION NUMBER: 2001-07-16
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565

PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 236
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-838-236

Query Match 4.1% Score 7: DB 9: Length 350;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AAAAAL 14
| | | | |
DB 338 AAAAAL 344

RESULT 37
US-09-907-613-236
; Sequence 236, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mathier, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090

PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-613-236

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 338 AAAAAA 344

RESULT 38

US-09-907-942-236
Sequence 236, Application US/09907942
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907, 942
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: PCT/US00/04414

PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143, 048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145, 698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146, 222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo sapiens
US-09-907-942-236

Query Match
Best Local Similarity 4.1%; Score 7; DB 9; Length 350;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AAAAAA 14
Db 338 AAAAAA 344

RESULT 39

US-09-904-820-236
Sequence 236, Application US/09904820
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James

```

APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-820-236

Query Match          4.1%; Score 7; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      8 AAAAAA 14
DB      338 AAAAAA 344

```

```

APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Guiney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavrin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 236
LENGTH: 350
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-859-236

Query Match          4.1%; Score 7; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      8 AAAAAA 14

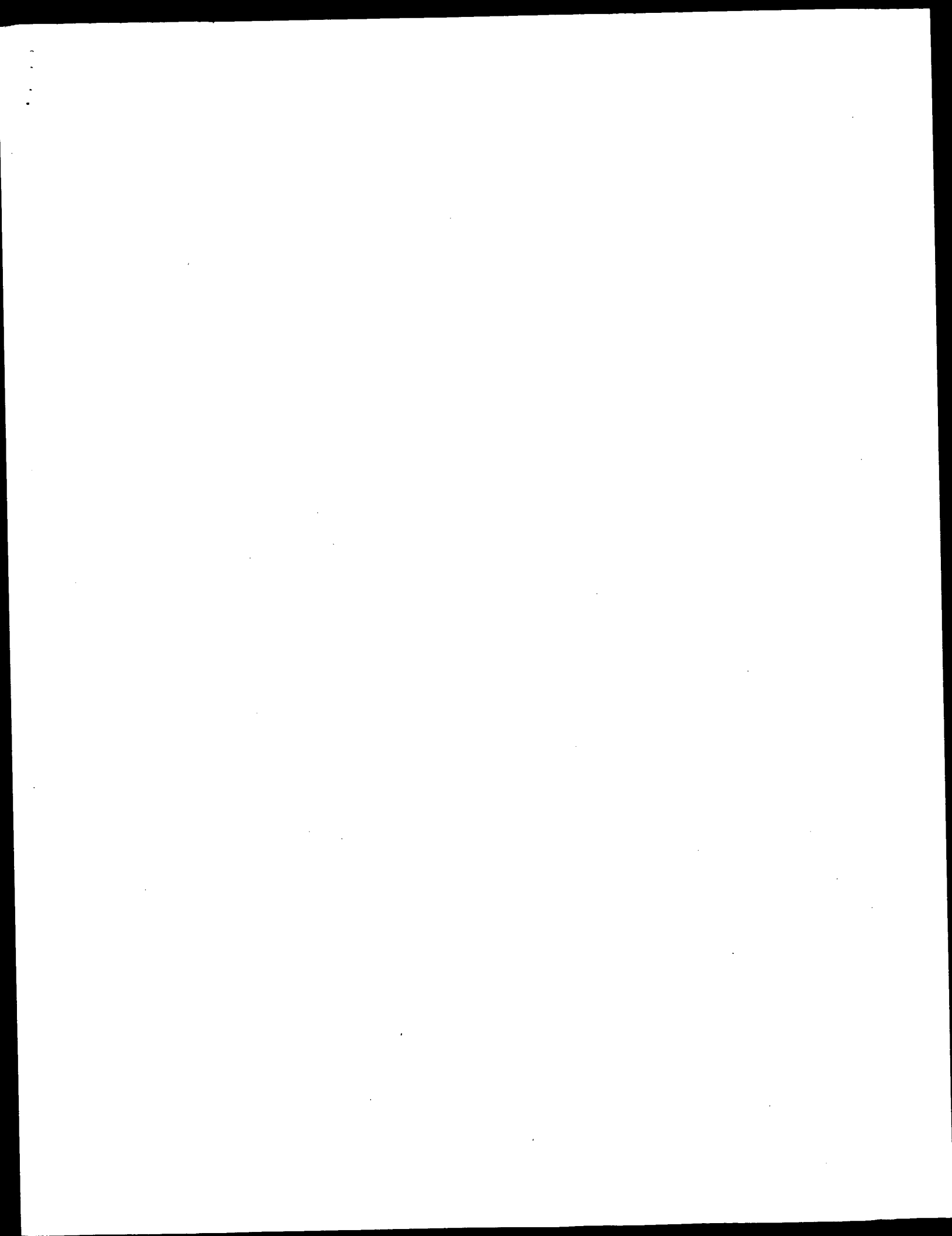
```


Mon Jul 7 08:55:22 2003

us-09-674-779b-2.oli.rapb

Db 338 |||||
AAAAAL 344

Search completed: July 6, 2003, 14:26:44
Job time : 24 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 13:40:40 ; Search time 35 Seconds
(without alignments)
203.827 Million cell updates/sec

Title: US-09-674-779B-2

Perfect score: 861

Sequence: 1 MMHIGTAAATAALSVLFEM.....IAFGTNEEASQNRRAELSY 172

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	316	36.7	166	1	PAL_PSEPU
2	285.5	33.2	173	1	PAL_ECOLI
3	264	30.7	153	1	PAL_HAEIN
4	258	30.0	150	1	PAL_PASMU
5	245	28.5	176	1	PAL_LBSPN
6	185.5	21.5	194	1	OMP_BORAV
7	176.5	20.5	353	1	OMP1_HAEIN
8	174	20.2	326	1	PORF_PSEFL
9	173.5	20.2	353	1	OMP2_HAEIN
10	171.5	19.9	350	1	PORF_PSEAE
11	165.5	19.2	359	1	OMP5_HAEIN
12	163.5	19.0	344	1	PORF_PSESY
13	154	17.9	219	1	YIND_ECOLI
14	151.5	17.6	236	1	OMP3_NEIGO
15	148.5	17.2	243	1	OMP4_ESCCE
16	148.5	17.2	346	1	OMP4_ECOLI
17	148.5	17.2	351	1	OMP4_SHIDY
18	144.5	16.8	242	1	OMP4_NEIMA
19	143	16.6	326	1	YB99_MYCTU
20	141	16.4	238	1	OMP4_CITFR
21	140.5	16.3	350	1	OMP4_SALTY
22	139.5	16.2	243	1	OMP4_ESCBE
23	139	16.1	241	1	OMP4_ESCBL
24	137.5	16.0	350	1	OMP4_ENTAE
25	136.5	15.9	344	1	OMP4_KLEPN
26	135.5	15.7	349	1	OMP4_BUCAI
27	127.5	14.8	293	1	MOTY_VIBAN
28	122.5	14.2	293	1	MOTY_VIBPA
29	122	14.2	359	1	OMP4_SERMA
30	117	13.6	417	1	TP50_TREPA
31	106.5	12.4	261	1	MOTB_BACSU
32	104	12.1	243	1	OMP4_SEROD
33	100	11.6	218	1	YTXE_BACME

34	98.5	11.4	150	1	YFIB_ECOLI	P07021	enterobacte
35	95.5	11.1	238	1	MOTB_TREPA	O07887	treponema p
36	87	10.1	257	1	MOTB_HELPY	P56427	helicobacte
37	86	10.0	242	1	YTXE_BACSU	P39064	bacillus su
38	83	9.6	551	1	G6P2_STRCO	Q92523	streptomyce
39	82	9.5	257	1	MOTB_HELPY	Q92129	helicobacte
40	81	9.4	639	1	COOS_RHOPU	P31896	rhodospiril
41	81	9.4	145	1	DTD_DEIRA	O84877	chlamydia t
42	80	9.3	361	1	TRM2_CHLPN	O92845	chlamydia p
43	80	9.3	378	1	DPB3_BACSU	P05649	bacillus su
44	80	9.3	378	1	SAHH_THEMEA	O51933	thermotoga
45	79	9.2	404	1	G6P1_STRCO	O88015	streptomyce
46	78.5	9.1	550	1	HE_PARLI	P22757	paracentrot
47	78.5	9.1	902	1	ITR1_PIG	O29052	sus scrofa
48	78	9.1	307	1	ARCC_HALNI	O48293	halobacteri
49	78	9.1	307	1	HMEN_DROVI	P09145	drosophila
50	76	8.8	584	1	ITR3_HUMAN	O06033	homo sapien
51	76	8.8	886	1	ITR3_MESAU	P97280	mesocricetu
52	76	8.8	886	1	ITR3_MOUSE	O61704	mus musculu
53	76	8.8	886	1	VTL2_RAT	O91151	rattus norv
54	75.5	8.8	224	1	ITR1_MOUSE	O61702	mus musculu
55	75.5	8.8	907	1	ITR1_MOUSE	O61702	mus musculu
56	75.5	8.8	3649	1	ACYS_NOCIA	P27743	nocardia la
57	75	8.7	399	1	TCR1_ECOLI	P02982	eschericchia
58	74.5	8.7	399	1	PSN1_PROME	P12881	drosophila
59	74.5	8.7	928	1	PMF9_CHLPN	O92398	chlamydia p
60	74	8.6	360	1	Y421_METUA	O57864	methanococc
61	74	8.6	360	1	YLO5_SULSO	P95879	sulfolobus
62	74	8.6	887	1	ITR3_RAT	O63416	rattus norv
63	73.5	8.5	330	1	IAFU_VIBPA	O03478	vibrio para
64	73.5	8.5	928	1	DEP1_ECOLI	P00582	eschericchia
65	73.5	8.5	1707	1	194K_TREVS	P05080	tobacco rat
66	73	8.5	460	1	VATB_METBA	O60187	methanosarc
67	73	8.5	535	1	VANI_YEAST	P23642	saccharomyc
68	72.5	8.4	351	1	COBT_MYCLE	O32953	mycobacteri
69	72.5	8.4	384	1	PROB_XYLEA	O99end	xyliella fas
70	72.5	8.4	460	1	VATB_METBA	P22663	methanosarc
71	72.5	8.4	521	1	CP11_SPAU	O42457	spatius aura
72	72.5	8.4	2431	1	POLN_SFV	P08411	semliki for
73	72	8.4	309	1	MOTB_SALTY	P55892	salmonella
74	72	8.4	320	1	Y054_MYCPN	P75049	mycoplasma
75	72	8.4	347	1	PYRC_ECO57	O83808	eschericchia
76	71.5	8.3	979	1	P115_MYCR	P41508	mycoplasma
77	71.5	8.3	1140	1	YW66_YEAST	O04883	saccharomyc
78	71	8.2	261	1	LPFR_MYCTU	P71798	mycobacteri
79	71	8.2	347	1	PYRC_ECOLI	P05020	eschericchia
80	71	8.2	362	1	PHEA_NEIGO	O92hy3	neisseria g
81	71	8.2	473	1	CYSN_BUCAI	P57498	buchnera ap
82	71	8.2	540	1	CALX_HELTO	O39994	heliobacter
83	71	8.2	606	1	PRIM_MYXXA	O50070	myxococcus
84	70.5	8.2	126	1	P15_MOUSE	P11031	mus musculu
85	70.5	8.2	416	1	PYRC_DEIRA	O92836	caenorhabdi
86	70.5	8.2	441	1	IF5_CAEEL	P06692	zymomonas m
87	70.5	8.2	552	1	HMEN_DROME	P53959	saccharomyc
88	70.5	8.2	568	1	DCPY_ZYMO	O33251	mycobacteri
89	70.5	8.2	839	1	YNEI_YEAST	O87601	pseudomonas
90	70	8.1	189	1	USG_PSEAE	P37604	salmonella
91	70	8.1	336	1	LPK_MYCTU	P46584	cyanoophora
92	70	8.1	390	1	DACD_SALTY	O94829	cyanoophora
93	70	8.1	501	1	YNS7_CONGL	O94829	cyanoophora
94	69.5	8.1	321	1	YC39_CAYPA	O50790	pseudomonas
95	69.5	8.1	351	1	HI81_PSEAE	P75563	mycoplasma
96	69.5	8.1	402	1	DP3B_MYCTU	P92778	mesocricetu
97	69.5	8.1	805	1	SYFB_MYCPN	P08300	rhodobacter
98	69.5	8.1	914	1	ITR1_MESAU		
99	69	8.0	191	1	UCR1_RHOCA		
100	69	8.0	401	1	NIRS_ENTAG		

RESULT 1

ALIGNMENTS

PAL_PSEPU STANDARD: PRT: 166 AA.
 ID PAL_PSEPU
 AC P43036;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL OR PAL1.
 OS Pseudomonas putida.
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 CC NCBI_TaxID=303;
 CC (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=mt-2;
 RX MEDLINE=96198174; PubMed=8626299;
 RA Rodriguez-Herva J.J., Ramos-Gonzalez M.I., Ramos J.L.;
 RT "The Pseudomonas putida peptidoglycan-associated outer membrane
 lipoprotein is involved in maintenance of the integrity of the cell
 cell envelope."
 RL J. Bacteriol. 178:1699-1706(1996).
 CC -1- FUNCTION: VERY STRONGLY ASSOCIATED WITH THE PEPTIDOGLYCAN (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.
 CC -1- SIMILARITY: TO OTHER PAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X74218; CAA52294.1;
 CC InterPro: IPR001145; Bac_OmpA.
 CC Pfam: PF00691; OmpA; 1.
 CC PRINTS: PD01021; OMPADOMAIN.
 CC PRODOM: PD000930; Bac_OmpA; 1.
 CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC DR PROSITE: PS01068; OMPA; 1.
 CC Outer membrane; Signal; Lipoprotein.
 CC FT SIGNAL 1 21 BY SIMILARITY.
 CC FT CHAIN 22 166 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
 CC FT LIPID 22 22 N-ACYL DIGLYCERIDE (BY SIMILARITY).
 CC FT DOMAIN 98 142 OMPA-LIKE.
 CC SQ SEQUENCE 166 AA; 17833 MW; 17353181435E5AC1 CRC64;
 Query Match 36.7%; Score 316; DB 1; Length 166;
 Best Local Similarity 43.2%; Pred. No. 9, 8e-22;
 Matches 73; Conservative 34; Mismatches 46; Indels 16; Gaps 6;
 YQ 6 QIAAAALSVLTFMTCANKSTISQ---VMVAPNAPPTGYTYGVAPVINDDETVKA 61
 DB 8 KFAALALMAV---AVGCCSSKGDNMCEGAADPNNGYAN---TGA---VDGSISEEA 57
 YQ 62 LASKLPVYEDFDEPKQAALIDEQAQFLTTQTAFLVAVAGHTDERGSRREYMSLG 121
 DB 58 -ALRAITTFEFDSSLPKPEARALDVHAKDKANGN-RVYLEGNTDEGTREYMAALG 115
 YQ 122 ERRVAVRNTLKGINKQASVEIISFGERRPIAFGNEEAWSONRAEL 170
 DB 116 ERRRAKAVQRYLVIGVSPADLELVSGEERPAVATGNDQESMAONRREL 164

DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL OR EXCC OR B0741 OR Z0909 OR ECS0776.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_TaxID=562, 83334;
 CC (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=87133578; PubMed=3545827;
 RA Chen R., Henning U.;
 RT "Nucleotide sequence of the gene for the peptidoglycan-associated
 lipoprotein of Escherichia coli K12."
 RL Eur. J. Biochem. 163:73-77(1987).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=92244043; PubMed=1574003;
 RA Lazzaroni J.-C., Portallier R.;
 RT "The exc gene of Escherichia coli K-12 required for cell envelope
 integrity encodes the peptidoglycan-associated lipoprotein (PAL).";
 RL Mol. Microbiol. 6:735-742(1992).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G., Ili, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizubuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horikuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., Ili, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533(2001).
 RN [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohsudo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Saekawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RN SEQUENCE OF 1-30 FROM N.A.
 RC STRAIN=JM105;
 RX MEDLINE=90078104; PubMed=2687247;

RA Levengood S.K., Webster R.E.;
 RT "Nucleotide sequences of the *tolA* and *tolB* genes and localization of
 RT their products, components of a multistep translocation system in
 RT *Escherichia coli*.";
 RL J. Bacteriol. 171:6600-6609(1989).
 RN [8]
 RP CRYSTALLIZATION.
 RX MEDLINE-21111468; PubMed-11173492;
 RA Abergel C., Walburger A., Chénayesse S., Lazdunski C.;
 RT "Crystallization and preliminary crystallographic study of the
 RT peptidoglycan-associated lipoprotein from *Escherichia coli*.";
 RL Acta Crystallogr. D 57:317-319(2001).
 CC -FUNCTION: Thought to play a role in bacterial envelope integrity.
 CC Very strongly associated with the peptidoglycan.
 CC -SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.
 CC -SIMILARITY: TO OTHER PAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X05123; CAA28771.1; -
 DR EMBL: X65796; CAA4673.1; -
 DR EMBL: AE000177; AAC73835.1; -
 DR EMBL: D90713; BAA35407.1; -
 DR EMBL: AE002522; AAG55077.1; -
 DR EMBL: AF002553; BAB34199.1; -
 DR EMBL: M28232; -; NOT_ANNOTATED_CDS.
 DR PIR: A27534; LPECPG.
 DR PIR: S20547; S20547.
 DR BioGene: E010684; pal.
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01068; OMPA; 1.
 DR Outer membrane; Signal; Lipoprotein; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 173 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
 FT LIPID 22 22 N-ACYL DIGLYCERIDE.
 FT DOMAIN 105 149 OMPA-LIKE.
 SQ SEQUENCE 173 AA; 18824 MW; 449F959C0274430 CRC64;
 Query Match 33.2%; Score 285.5; DB 1; Length 173;
 Best Local Similarity 37.6%; Pred. No. 5.9e-19;
 Matches 67; Conservative 33; Mismatches 67; Indels 11; Gaps 3;
 OY 1 MHLHIQIAAAAAAIVLFTFMTCAGKSTSQVAVNAATPGY---TGVIYVGAFLVND 56
 Db 1 MQLNVYKGLMTALVPMALVACSSNKNKSN----DSEGMGAGCTGANGGNGMSSE 55
 OY 57 ETVKALASKLP--SLVYEDFSDSEIKPQAAAILDFOAFLITNORAVLVAGHDERSSR 114
 Db 56 EQARLQMOQLQONNIVYEDLDKXDIRSDEAQMLDHANFLNSPSKYKTVESHADERGTP 115
 OY 115 EYNNMISGERRAVAVNNYILGKGINQASVEIISFGEERPIATGCTNEANSQNRRLASY 172
 Db 116 EYNNISLGERRAVAVNNYILGKGINQASVEIISFGEERPIATGCTNEANSQNRRLASY 173
 RESULT 3
 PAL_HAEIN STANDARD; PRT; 153 AA.
 AC P10324;
 DT 01-MAR-1989 (Rel. 10; Created)
 DT 01-MAR-1989 (Rel. 10; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)

DE Outer membrane protein P6 precursor (OMP P6) (15 kDa peptidoglycan-
 DE associated lipoprotein) (PC protein).
 GN PAL OR OMP6 OR HI0381.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxId=727;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE-88115138; PubMed-2828309;
 RA Delch R.A., Metcalf B.J., Finn C.W., Farley J.E., Green B.A.;
 RT "Cloning of genes encoding a 15,000-dalton peptidoglycan-associated
 RT outer membrane lipoprotein and an antigenically related 15,000-dalton
 RT protein from *Haemophilus influenzae*.";
 RL J. Bacteriol. 170:489-498(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-88085463; PubMed-3257200;
 RA Nelson M.B., Apicella M.A., Murphy T.F., Vankeulen H., Spottla L.D.,
 RA Rekosh D.;
 RT "Cloning and sequencing of *Haemophilus influenzae* outer membrane
 RT protein pe.";
 RL Infect. Immun. 56:128-134(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Rd / KW20 / ATCC 51907;
 RX MEDLINE-95350630; PubMed-7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of *Haemophilus influenzae*
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
 CC anchor.
 CC -SIMILARITY: TO OTHER PAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M19391; AAA24994.1; -
 DR EMBL: M18878; AAA24940.1; -
 DR EMBL: U32722; AAC22039.1; -
 DR PIR: A27558; A27558.
 DR PIR: A28543; A28543.
 DR TIGR: H10381; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01068; OMPA; 1.
 DR Outer membrane; Signal; Lipoprotein; Complete proteome.
 FT SIGNAL 1 19
 FT CHAIN 20 153 OUTER MEMBRANE PROTEIN P6.
 FT LIPID 20 20 N-ACYL DIGLYCERIDE.
 FT DOMAIN 85 129 OMPA-LIKE.
 SQ SEQUENCE 153 AA; 16108 MW; 3DF358122EF17A11 CRC64;
 Query Match 30.7%; Score 264; DB 1; Length 153;
 Best Local Similarity 49.5%; Pred. No. 4.4e-17;
 Matches 51; Conservative 20; Mismatches 32; Indels 0; Gaps 0;

QY	70	VVFEDPSDEIKRQAAMILDEDAOFITTMOTARVLVACHTDERGSRKSNMISLGERRAVAVR	129
Dd	51	VVFEGDKVIDIGTEIYYOILLDAAIAALYNATPAKVALYVESGNIDKERPEPTINIALGORADAYK	110
QY	130	NYLLKGAINQASVELLIISFGEERPIAFGTNEEANSOUNRELA5Y 172	
		: : :: : : :: : :	
		111 GYLAKGVADKLGTVSYGEEKPAVLGHDEAAYSKNBRAVLAV 153	
RESULT 4			
PAL_PASMO	STANDARD:	PRT:	150 AA.
ID	PAL_PASMO		
AC	051886;		
DT	16-OCT-2001 (Rel. 40; Created)		
DT	16-OCT-2001 (Rel. 40; Last sequence update)		
DT	16-OCT-2001 (Rel. 40; Last annotation update)		
DE	Outer membrane protein P6 precursor (OMP P6) (P6-like) (Peptidoglycan-associated lipoprotein).		
GN	PAL OR PM0956.		
OS	Pasteurella multocida.		
OC	Bacteriella Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Pasteurella.		
OX	NCBI_TaxID=747;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=116;		
RX	MEDLINE=95172751; PubMed=7868272;		
RA	Kasten R.W., Hansen L.M., Hinojosa J., Bieber D., Ruehl W.W.,		
RA	Hirsh D.C.;		
RT	"Pasteurella multocida produces a protein with homology to the P6		
RT	outer membrane protein of Haemophilus influenzae.";		
RL	Infect. Immun. 63:989-993(1995).		
RM	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Pm70;		
RX	MEDLINE=21145866; PubMed=11248100;		
RA	May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;		
RT	"Complete genomic sequence of Pasteurella multocida Pm70."		
RL	Proc Natl Acad Sci U S A. 98:3460-3465(2001).		
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID ANCHOR (By similarity).		
CC	-1- SIMILARITY: TO OTHER PAL PROTEINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL: U16849; AAA79373.1; -		
DR	EMBL: AE006136; AAK03050.1; -		
DR	InterPro: IPR001145; Bac_OmpA.		
DR	Pfam: PF00691; OMPA; 1.		
DR	PRINTS: PR01021; OMPADOMAIN.		
DR	ProDom: PD000930; Bac_OmpA; 1.		
DR	ProSITE: PS00013; PROKAR_LIPOPROTEIN_1.		
KW	PROSITE; PS01068; OMPA; 1.		
FT	Outer membrane; Signal; Lipoprotein; Complete proteome.		
FT	SIGNAL	1	19
FT	CHAIN	20	150
FT	LIPID	20	20
FT	DOMAIN	82	126
SO	SEQUENCE	150 AA; 16213 MW; 170E7B13D2EBEDC CRC64;	
	Query Match	30.0%; Score 258; DB 1; Length 150;	
	Best Local Similarity	48.5%; Pred. No. 1.5e-16;	
	Matches	50; Conservative 19; Mismatches 34; Indels 0; Gaps 0;	
QY	70	VVFEDPSDEIKRQAAMILDEDAOFITTMOTARVLVACHTDERGSRKSNMISLGERRAVAVR	129
		: : :: : : :: : :	
		: :: : : :: : : :: : :	

Db 48 VYGEFDKYNIEGVEVQILDAHAFLNLPATKVYVEGNTEBGRPEXNALGQRADAVK 107

Oy 130 NYLKGKINGMASVEITISFGERRPIAFCTNEAMSONRAELSY 172
: : : : :
Db 108 HTLSAKGVAGQVSTVSYSGERPAVLGHDEAAYSKNRRAYLAY 150

RESULT 5
PAL_LEGPN STANDARD; PRT: 176 AA.

ID PAL_LEGPN
AC P26493;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Peptidoglycan-associated lipoprotein precursor (19 kDa surface antigen) (PPL).
GN PAL OR PPVA.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OX Legionellaceae; Legionella.
NCBI_TaxID=446;
[1]
SEQUENCE FROM N.A.
RP STRAIN-AA100 / Serogroup 1;
RC Enleberg N.C., Howe D.C., Rogers J.E., Arroyo J., Eisenstein B.I.;
RA Engleberg N.C., Howe D.C., Rogers J.E., Arroyo J., Eisenstein B.I.;
RT "Characterization of a Legionella pneumophila gene encoding a lipoprotein antigen."
RL Mol. Microbiol. 5:2021-2029(1991).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=91310296; PubMed=1855972;
RX Ludwig B., Schmid A., Marre R., Hacker J.;
SA "Cloning, genetic analysis, and nucleotide sequence of a determinant coding for a 19-kilodalton peptidoglycan-associated protein (Ppl) of Legionella pneumophila."
RT Infect. Immun. 59:2515-2521(1991).
RL -I- FUNCTION: VERY STRONGLY ASSOCIATED WITH THE PEPTIDOGLYCAN.
CC -I- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid anchor.
CC -I SIMILARITY: TO OTHER PAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

DR EMBL: X60543; CAA43033.1; -.
DR PIR: A60337; A60337.
DR PIR: S16631; S16631.
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_OmpA; 1.
DR PROSITE: PS00013; PROKAR_LIPOCOREIN; 1.
DR PROSITE: PS01068; OMPA; 1.
KW Lipoprotein; Outer membrane; Antigen; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 1 21 PEPTIDOGLYCAN-ASSOCIATED LIPOPROTEIN.
FT LIPID 22 22 N-ACYL DIGLYCERIDE (PROBABLE).
FT DOMAIN 104 148 OMPA-LIKE.
SO SEQUENCE 176 AA; 18911 MW; 7D9C3BBECBE621DB CRC64;

Query Match 28.5%; Score 245; DB 1; Length 176;
Best Local Similarity 36.0%; Pred. No. 2.7e-15;
Matches 58; Conservative 30; Mismatches 53; Indels 20; Gaps 3;

Oy 29 SQWVAPAPRGYGVIVTGAPLVNDNEFKAL-----ASKLP-----SLVY 71
: : : : :
Db 15 SAVLVASCKTPGSA---DGGAANDGATAGOGLOGMTHFAGQEGESYTTOAPHNQYL 71

```

OY 72 FDPSDEIKPQAAAILDEAOFLTTNQTARVLAHTDEGRSEYMSLGERRAVAVRNY 131
DB 72 FAYDSTLASKYLSVNAQAEYKLTHTGPARVMIGHDDEKRSREYNALGERRADTVAET 131
OY 132 LIGKGINOASVEIITSGFERPIAFGTNEAMSQNRRAELSY 172
DB 132 LRMAGVSRQOIRVVSYGKERBANYGHDEASHAQNRRVEFTY 172

RESULT 6
OMPA_BORAV STANDARD; PRT; 194 AA.
ID OMPA_BORAV
AC 005146;
DT 01-OCT-1994 (Rel. 30, Created)
DR 01-OCT-1994 (Rel. 30, Last sequence update)
DE 01-FEB-1995 (Rel. 31, Last annotation update)
DE Outer membrane protein A precursor.
GN OMPA.
OS Bordetella avium.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_Taxid=521;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=197;
RA MEDLINE=93077456; PubMed=1447140;
RA Gentry-Weeks C.R., Hultsch A.-L., Kelly S.M., Kelth J.M.,
RA Curtiss R. III;
RT "Cloning and sequencing of a gene encoding a 21-kilodalton outer
RT membrane protein from Bordetella avium and expression of the gene in
RT Salmoneella typhimurium.";
RL J. Bacteriol. 174:7729-7742(1992).
CC -1- FUNCTION: STRUCTURAL PROTEIN THAT MAY PROTECT THE INTEGRITY OF
CC THE BACTERIUM.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M96550; AAA22979.1; -.
DR PIR: A45275; A45275.
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR PRODOM: PD000930; Bac_OmpA; 1.
DR PROSITE: PS01068; OMPA; 1.
KW Outer membrane; Transmembrane; Porin; Antigen; Signal.
FT SIGNAL 1 24
FT CHAIN 25 194 OUTER MEMBRANE PROTEIN A.
FT DOMAIN 54 76 PRO-RICH.
FT DOMAIN 121 165 OMPA-LIKE.
SQ SEQUENCE 194 AA; 21115 MW; 1A25F2A6367DAE7A CRC64;

Query Match 21.5%; Score 185.5; DB 1; Length 194;
Best Local Similarity 30.7%; Pred. No. 7.3e-10;
Matches 47; Conservative 22; Mismatches 63; Indels 21; Gaps 2;

OY 18 TFMTCGANKSTSCVMAVNPATGYTGVIYTGVAFLVNDDETVAKALSKPLSLVFPDSD 77
DB 56 TGIPGCDGVVQAQPEKAPM-----AAKVFNADT-----FDFDKS 94
OY 78 EIRPQAAIILDEAOFLTTNQTARVLAHTDEGRSEYMSLGERRAVAVRNYLIGKCI 137
DB 95 TLKPEGRQLDDVQAQARAIDLETTIIVAGTDSIGTEAYNMKLSERRASVAVRYLAVSKI 154
OY 138 NQASVELISFGEERPIAFGTNEAMSQNRRAEL 170

```

```

DB 155 DPNRITTEGKGLNPLASNKTAGRAANRRVEL 187.
OY 155 DPNRITTEGKGLNPLASNKTAGRAANRRVEL 187.

RESULT 7
OM51_HAEIN STANDARD; PRT; 353 AA.
ID OM51_HAEIN
AC P43840;
DT 01-NOV-1995 (Rel. 32, Created)
DR 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein P5 precursor (OMP P5).
GN OMPA OR OMP5 OR H1164.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Gnehm C.L., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd".
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32796; AAC2819.1; -.
DR HSP: P02934; IQDP.
DR TIGR: H1164; -.
DR InterPro: IPR001145; Bac_OmpA.
DR InterPro: IPR000498; OMPA_tmrm.
DR Pfam: PF00691; OmpA; 1.
DR Pfam: PF01389; OmpA_membrane; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR PRODOM: PD000930; Bac_OmpA; 1.
DR PROSITE: PS01068; OMPA; 1.
KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 353
FT DISULEID 326 338 BY SIMILARITY.
FT DOMAIN 272 316 OUTER MEMBRANE PROTEIN P5.
FT DOMAIN 272 316 OMPA-LIKE.
SQ SEQUENCE 353 AA; 37743 MW; 64ACB3E7BFF96B39 CRC64;

Query Match 20.5%; Score 176.5; DB 1; Length 353;
Best Local Similarity 30.6%; Pred. No. 9.8e-09;
Matches 48; Conservative 27; Mismatches 61; Indels 21; Gaps 4;

OY 35 PNAPTGY-----TGVIY---TGVAFLVNDDETVAKALSKPLSLVFPDSDIIRPQA 83
DB 193 PNTALNYPMPWIGSINAGISYRFQGAAPVVAPEVSKTSS-LNSDVTFAAGKANLKPQA 251
OY 84 AAIDDEAOFLTTNQTARVLAHTDEGRSEYMSLGERRAVAVRNYLIGKGINQASVE 143
DB 252 QATIDSTIGEMSOVSAKAVAVAGYTDRIQSDAFNVKLSQERADSVAVYFAKGVADAIS 311

```

OY 144 IISGGERPIAFGTNEE-----AMSONRRAELS 171
DB 312 ATGYKAMPVTGATCDGVKGRKALLACFADPRRVEIA 348

RESULT 8

PORT_PSEFL

STANDARD; PRT; 326 AA.

AC P37726;

01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane porin F precursor (Root adhesin).

OPRF.

OS Pseudomonas fluorescens.

Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

NCBI_TaxID=294;

OX NCBI_TaxID=294;

SEQUENCE FROM N.A., AND SEQUENCE OF 25-74 AND 238-258.

RC STRAIN-OE 28.3;

MEDLINE-92167971; PubMed-1538702;

RA de Mot R., Proost P., Van Damme J., Vanderleyden J.;

"Homology of the root adhesin of Pseudomonas fluorescens OE 28.3 with

RT porin F of P. aeruginosa and P. syringae.";

Mol. Gen. Genet. 231:489-493(1992).

RL Mol. Gen. Genet. 231:489-493(1992).

-1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.

CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO OTHER

PSEUDOMONAS SPECIES OPRF.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

EMBL: AF115334; AAD45981.1;

DR InterPro: IPR001145; Bac_OmpA.

Pfam: PF00691; OmpA.1.

DR PRINTS: PRO1021; OMPADOMAIN.

DR PRODOM: PD000930; Bac_OmpA.1.

DR PROSITE: PS01068; OMPA.1.

KW Outer membrane; Transmembrane; Porin; Signal; Repeat.

FT SIGNAL

1 24

FT CHAIN

25 326 OUTER MEMBRANE PORIN F.

FT DOMAIN

190 205 8 X 2 AA TANDEN REPEATS OF X-P.

FT REPEAT

190 191 1.

FT REPEAT

192 193 2.

FT REPEAT

194 195 3.

FT REPEAT

196 197 4.

FT REPEAT

198 199 5.

FT REPEAT

200 201 6.

FT REPEAT

202 203 7.

FT REPEAT

204 205 8.

FT DOMAIN

253 297 OMPA-LIKE.

SO SEQUENCE

326 AA; 34482 MW; 4C183F56A84D2A76 CRC64;

Query Match 20.2%; Score 174; DB 1; Length 326;
Best Local Similarity 36.3%; Pred. No. 1.5e-08;
Matches 37; Conservative 14; Mismatches 51; Indels 0; Gaps 0;

OY 70 VYFDFSDSEIRKPOAAIIDEQOAFITNOTARVLVAGHTDERSREYNSLGERRAVAVR 129
DB 219 VKFDFDKSVKPNISGVKYNLADFMAYQPATINVEYAGHTDERSIGDPAITNOKLSQRADRVK 278

OY 130 NYLKGGINOASVELISGGERPIAFGTNEEAMSONRRAELS 171
DB 279 QVLVVDGVAAPSRITAVGGEERPAVDANTEAGRAVNRREVS 320

RESULT 9
OM52_HAEIN STANDARD; PRT; 353 AA.

AC P38368;

01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

01-NOV-1995 (Rel. 32, Last annotation update)

DE Outer membrane protein p5 precursor (OMP p5).

OMP p5.

GN OMPA OR OMP5.

Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

Haemophilus.

OC NCBI_TaxID=727;

OX NCBI_TaxID=727;

SEQUENCE FROM N.A., AND SEQUENCE OF 22-34.

RC STRAIN-1613 / Serotype B;

MEDLINE-93366472; PubMed-8359929;

RA Munson R.S. Jr., Grass S., West R.;

"Molecular cloning and sequence of the gene for outer membrane

RT protein p5 of Haemophilus influenzae.";

Infect. Immun. 61:4017-4020(1993).

RL Infect. Immun. 61:4017-4020(1993).

-1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.

CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

EMBL: I20309; AAA03346.1;

DR HSP: P02934; IOLP.

InterPro: IPR001145; Bac_OmpA.

DR InterPro: IPR000498; OmpA_tmam.

Pfam: PF01389; OmpA.1.

DR PRINTS: PRO1021; OMPADOMAIN.

DR PRODOM: PD000930; Bac_OmpA.1.

DR PROSITE: PS01068; OMPA.1.

KW Outer membrane; Transmembrane; Porin; Signal.

FT SIGNAL

1 21

FT CHAIN

22 353 OUTER MEMBRANE PROTEIN P5.

FT DISULFID

326 338 BY SIMILARITY.

FT DOMAIN

272 316 OMPA-LIKE.

SO SEQUENCE

353 AA; 37594 MW; E58A659E786D0F7 CRC64;

Query Match 20.2%; Score 173.5; DB 1; Length 353;
Best Local Similarity 30.6%; Pred. No. 1.8e-08;
Matches 48; Conservative 26; Mismatches 62; Indels 21; Gaps 4;

OY 35 PNAPTYG-----TGVIY---TGVAFLVNDDETVKALASKLPVLYFDFSDSEIRKPOA 83
DB 193 PTAIYNYNWISINAGISYRFGQGAAPVVAPEVYSKTFE-LNSDVFARFGKANLKPOA 251

OY 84 AAILDEQOAFITNOTARVLVAGHTDERSREYNSLGERRAVAVRNTLLKGINOASVE 143
DB 252 QATLDSIYEMSOVSASAKVAVAGYTDRIQSDSADPNVKLSQERADSVANYFAKVAADALS 311

OY 144 IISGGERPIAFGTNEE-----AMSONRRAELS 171
DB 312 ATGYKAMPVTGATCDGVKGRKALLACFADPRRVEIA 348

RESULT 10

PORT_PSEAE

STANDARD; PRT; 350 AA.

AC P13794;

01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane porin F precursor.

GN OPRF OR PA1777.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88086862; PubMed=2447060;
 RA Ducheine M., Schweizer A., Lottspeich F., Krauss G., Margot M.,
 RA Vogel K., von Specht B.-U., Domdey H.;
 RT "Sequence and transcriptional start site of the Pseudomonas
 aeruginosa outer membrane porin protein F gene.";
 RL J. Bacteriol. 170:155-162(1988).
 [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 15692 / PA01;
 RC MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garner R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lattig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 CC -1- FUNCTION: HAS PORIN ACTIVITY, FORMING SMALL WATER-FILLED CHANNELS.
 CC ALSO HAS A STRUCTURAL ROLE IN DETERMINING CELL SHAPE AND ABILITY
 CC TO GROW IN LOW-OSMOLARITY MEDIUM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- PTM: TWO DISULFIDE BONDS ARE PRESENT.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO OTHER
 CC PSEUDOMONAS SPECIES OPRF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF027290; AAD11568.1; -;
 DR EMBL: M18975; AAA25973.1; -;
 DR EMBL: A01477; CAA00171.1; -;
 DR EMBL: AE004603; AAG05166.1; -;
 DR PIR: S39156; S39156.
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; 1.
 KW Outer membrane; Porin; Transmembrane; Signal; Repeat;
 KW Complete proteome.
 FT SIGNAL 1 24
 FT CHAIN 25 350 OUTER MEMBRANE PORIN F.
 FT DOMAIN 190 195 3 X 2 AA TANDEM REPEATS OF X-P.
 FT REPEAT 190 191 1.
 FT REPEAT 192 193 2.
 FT REPEAT 194 195 3.
 FT DOMAIN 276 321 OMPA-LIKE.
 SO SEQUENCE 350 AA; 37639 MW; D1F758F076874383 CRC64;
 Query Match 19.9%; Score 171.5; DB 1; Length 350;
 Best Local Similarity 37.6%; Pred. No. 2.7e-08;
 Matches 38; Conservative 15; Mismatches 47; Indels 1; Gaps 1;
 QY 70 VFEDSDSEIKPQAAILDEQAQFLTTNQTARVAVAGTDERGSRREYMSIGERRAVAVR 129
 Db 242 VFEDSDSEIKPQAAILDEQAQFLTTNQTARVAVAGTDERGSRREYMSIGERRAVAVR 129
 QY 130 NYLLDKGINQASVELISGGERPIAFGTNEASQNRRAE 169
 Db 130 NYLLDKGINQASVELISGGERPIAFGTNEASQNRRAE 169

Db 302 DVLVNEYGVEGGRVNAVGYGESRPPVADNATAGRAIRRYVE 342
 RESULT 11
 ID OM53_HAEIN STANDARD; PRT; 359 AA.
 AC P45996;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane protein P5 precursor (OMP P5) (Fimbria).
 GN OMPA OR OMP5.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX STRAIN=NH1 1128;
 RC MEDLINE=94222575; PubMed=7909539;
 RA Strakova T., Kolattukudy P.E., Murwin D., Billy J., Leake E.,
 RA Lim D., Demaria T., Bakaletz L.;
 RT "Role of fimbriae expressed by nontypeable Haemophilus influenzae in
 RT pathogenesis of and protection against otitis media and relatedness
 RT of the fimbria subunit to outer membrane protein A.";
 RL Infect. Immun. 62:2002-2020(1994).
 CC -1- FUNCTION: ACTS AS A FIMBRIAE SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: I08448; AAA24959.1; -;
 DR HSSP: P02934; 1BXW.
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OmpA_tmem.
 DR Pfam: PF00691; OmpA; 1.
 DR Pfam: PF01389; OmpA_membrane; 1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; FALSE_NEG.
 KW Outer membrane; Transmembrane; Porin; Signal; Fimbria.
 FT SIGNAL 1 21
 FT CHAIN 22 359 OUTER MEMBRANE PROTEIN P5.
 FT DISULFID 332 344 BY SIMILARITY.
 FT DOMAIN 278 322 OMPA-LIKE.
 SO SEQUENCE 359 AA; 38340 MW; 576B1C5B4818C37 CRC64;
 Query Match 19.2%; Score 165.5; DB 1; Length 359;
 Best Local Similarity 29.9%; Pred. No. 9.9e-08;
 Matches 47; Conservative 26; Mismatches 63; Indels 21; Gaps 4;
 QY 35 PNAFTGY-----TGVIY---TGAVALVDNDETAKALSKLPISLVYEDFSDSEIKPQA 83
 Db 199 PNTAINTNPMWIGICINAGISYRFGGGEAPVVAADAEVMSKTPS-LNSDYTFAGKANLKPOA 257
 QY 84 AAILDEQAQFLTTNQTARVAVAGTDERGSRREYMSIGERRAVAVANNYLKGINQASVE 143
 Db 258 QATIDSYGEISQVSKKAVAGTNTNIGSDATNKLSDGRADSVANFYAKGVADAIS 317
 QY 144 IISGGERPIAFGTNEE-----AWSQNRRAEIS 171
 Db 318 ATGYGEANPYTGATCDQVKGKRALINCLAPDRRVEIA 354
 RESULT 12
 PORF_PSESX

ID	PORE_PSESY	STANDARD;	PRF:	344 AA.
AC	P22263;			
DT	01-AUG-1991 (Rel. 19,	Created)		
DT	01-AUG-1991 (Rel. 19,	Last sequence update)		
DT	15-JUL-1999 (Rel. 38,	Last annotation update)		
DE	Outer membrane porin F precursor.			
GN	OprF.			
OS	Pseudomonas syringae (pv. syringae).			
CC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCB1_taxonomy:321;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 19310;			
RX	MEDLINE=91100367; PubMed=1898935;			
RA	Ullstrom C.A., Steinhel R., Woodruff W., Steinbach S., Hancock R.E.W.;			
RT	"Conservation of the gene for outer membrane protein Oprf in the family Pseudomonadaceae: sequence of the Pseudomonas syringae oprf gene."			
RL	J. Bacteriol. 173:768-775(1991).			
CC	-1- FUNCTION: HAS PORIN ACTIVITY, FORMING SMALL WATER-FILLED CHANNELS.			
CC	-1- ALSO HAS A STRUCTURAL ROLE IN DETERMINING CELL SHAPE AND ABILITY TO GROW IN LOW-OSMOLARITY MEDIUM.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.			
CC	-1- PTM: TWO DISULFIDE BONDS ARE PRESENT.			
CC	-1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO OTHER PSEUDOMONAS SPECIES OPRE.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M55408; AAA25910.1; .			
DR	PIR: A39139; A39139.			
DR	InterPro: IPR001145; Bac_OmpA.			
DR	Pfam: PF00691; OmpA; 1.			
DR	PRINTS: PRO1021; OMPADOMAIN.			
DR	Prodor: PD000930; Bac.OmpA; 1.			
DR	PROSITE: PS01068; OMPA; 1.			
KW	Outer membrane; Transmembrane; Porin; Signal.			
FT	SIGNAL 1 24 BY SIMILARITY.			
FT	CHAIN 25 344 OUTER MEMBRANE PORIN F.			
FT	DOMAIN 270 315 OMPA-LIKE.			
SQ	SEQUENCE 344 AA; 36567 MW; 58463D9AD973D755 CRC64;			
Query Match 19.0%; Score 163.5; DB: 1; Length 344;				
Best Local Similarity 35.6%; Pred. No. 1.4e-07;				
Matches 36; Conservative 14; Mismatches 50; Indels 1; Gaps 1;				
QY	70 VFPPDSDELKPOAAALIDEQAOFLTNQTARVLVAGHTDGSGREYMMSLGERRAVVR 129 I : : : : :			
Dd	236 VKFPDFSKSVVPKNVSVDGIKLNDLPMDQQPQPTTVTEGHTSDVGDDATNQKISERRANVK 295 : : : : :			
QY	130 NYLLQG-GINDASYEIIISFGGERPIAGTGNEANSQNRRAE 169 : : : : : : : :			
Dd	296 QVLVNQYGVASRVNSGVGEKSKEPADMTAGRANVRKY 336 : : : : : : : :			
RESULT 13				
ID	XIAD_ECOLI	STANDARD;	PRF:	219 AA.
AC	P37665;			
DT	01-OCT-1994 (Rel. 30,	Created)		
DT	01-OCT-1994 (Rel. 30,	Last sequence update)		
DT	16-OCT-2001 (Rel. 40,	Last annotation update)		
DE	Hypothetical lipoprotein yiad precursor.			
GN	YIAD OR B3552.			
OS	Escherichia coli.			

DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Outer membrane protein P.III precursor (Gonococcal protein III)
DE (Pili).
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87139801; PubMed=3102671;
RA Gotschlich E.C., Seiff M., Blake M.S.;
RT "The DNA sequence of the structural gene of gonococcal protein III
RT and the flanking region containing a repetitive sequence. Homology of
RT protein III with enterobacterial Ompa proteins.";
RL J. Exp. Med. 165:471-482(1987).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- MISCELLANEOUS: Pili is closely associated with pili, but NOT ALL
CC OF THE PI MOLECULES IN THE GONOCOCCAL OUTER MEMBRANE.
CC -1- MISCELLANEOUS: A PORTION OF THE Pili IS EXPOSED TO THE SURFACE IN
CC INTACT GONOCOCCI; THE PROTEIN CAN BE LABELED WITH IODINE AND
CC REACTS WITH MABS.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY. STRONG, TO N.MENINGITIDIS
CC RMPM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05105; CAA28752.1; -;
DR PIR: A27894; A27894.
DR InterPro: IPR001145; Bac_Ompa.
DR Pfam: PF00691; Ompa.1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_Ompa.1.
DR PROSITE: PS01068; Ompa.1.
KW Outer membrane; Porin; Transmembrane; Antigen; Signal; Repeat.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 236 OUTER MEMBRANE PROTEIN P.III.
FT DOMAIN 69 76 4 X 2 AA TANDEM REPEATS OF X-P.
FT REPEAT 69 76 1.
FT REPEAT 71 72 2.
FT REPEAT 73 74 3.
FT REPEAT 75 76 4.
FT DOMAIN 131 175 OMPA-LIKE.
FT DISULFID 185 208 BY SIMILARITY.
SQ SEQUENCE 236 AA; 25540 MW; DAE9AAECA6FB199 CRC64;
Query Match 17.6%; Score 151.5; DB 1; Length 236;
Best Local Similarity 36.5%; Pred. No. 1.1e-06;
Matches 35; Conservative 13; Mismatches 43; Indels 5; Gaps 1;
QY 56 DETVATLASKLPSTLYVFPDSEIKPQAAALIDEOAQLTNTQARVAVAGHTDEQRSRE 115
DB 88 DETISLSAKTL-----FGFDKSLRAEQNDLKLAKLQRLSTNQSVRVEGHTTFPMSEK 142
QY 116 YNMSLGERRAVAVRNLYLGKGINOASVEIISFGEER 151
DB 143 YNQAISERRAVVANNILVSGVPASRISAVGLGESQ 178
RESULT 15
OMPA_ESCFE STANDARD; PRT; 243 AA.
AC P24747;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Outer membrane protein A (Outer membrane protein II) (Fragment).
GN OMPA.

OS Escherichia fergusonii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=564;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35471, ATCC 35469, and ATCC 35472;
RX MEDLINE=92065252; PubMed=1955870;
RA Lawrence J.G., Ochman H., Hartl D.L.;
RT "Molecular and evolutionary relationships among enteric bacteria";
RL J. Gen. Microbiol. 137:1911-1921(1991).
CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHASES. ALSO ACTS AS A PORIN
CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
CC SOLUTES (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAINS ATCC 35471
CC AND ATCC 35472.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63352; AAA24236.1; -;
DR EMBL: M63351; AAA24232.1; -;
DR EMBL: M63353; AAA24240.1; -;
DR HSP: P02934; 10JP.
DR InterPro: IPR001145; Bac_Ompa.
DR InterPro: IPR000498; Ompa_tmem.
DR Pfam: PF00691; Ompa.1.
DR ProDom: PD000930; Bac_Ompa.1.
DR PROSITE: PS01068; Ompa.1.
KW Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
FT NON_TER 1 1
FT TRANSMEM 1 8 POTENTIAL.
FT TRANSMEM 12 27 POTENTIAL.
FT TRANSMEM 48 59 POTENTIAL.
FT TRANSMEM 65 81 POTENTIAL.
FT TRANSMEM 87 98 POTENTIAL.
FT DOMAIN 107 114 4 X 2 AA TANDEM REPEATS OF A-P.
FT REPEAT 107 108 1.
FT REPEAT 109 110 2.
FT REPEAT 111 112 3.
FT REPEAT 113 114 4.
FT DOMAIN 163 207 OMPA-LIKE.
FT DISULFID 217 229 BY SIMILARITY.
FT VARIANT 39 39 E -> D (IN STRAIN ATCC 35469).
FT NON_TER 243 243
SQ SEQUENCE 243 AA; 26144 MW; B89F7FC29D42E3AB CRC64;
Query Match 17.2%; Score 148.5; DB 1; Length 243;
Best Local Similarity 31.6%; Pred. No. 2.1e-06;
Matches 42; Conservative 20; Mismatches 60; Indels 11; Gaps 2;
QY 49 VAPLVNDDEVKALAKLPSTLYVFPDSEIKPQAAALIDEOAQLTNTQARVAVAG 106
DB 106 VAPAPAPAPVQKHTLTLSVDVLFNFKATLRQGAALDQLXSQLSNIDPKDGSVYVLG 165
QY 107 HTDEGSRREYNMSLGERRAVAVRNLYLGKGINOASVEIISFGEERPIAGTNE----- 159
DB 166 YTRIGSDAYNQGLSERRAQSVYDYLISGIPADKISAGMGESNVTGTCDDNKKQRAA 225
QY 160 --EAMSONRAEL 170
DB 226 LIDCLAPDRRVEI 238

RESULT 16
OMPA_ECOLI STANDARD; PRT: 346 AA.
AC P02934;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (Outer membrane protein II*).
GN OMPA OR TOLG OR TUP OR CON OR B0957 OR Z1307 OR ECS1041.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=81053729; PubMed=6253901;
RA Beck E., Bremer E.;
RT "Nucleotide sequence of the gene ompa coding the outer membrane
RL protein II of Escherichia coli K-12.";
RN [2]
RP Nucleic Acids Res. 8:3011-3024(1980).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=81170587; PubMed=6260961;
RA Movva N.R., Nakamura K., Inouye M.;
RT "Gene structure of the ompa protein, a major surface protein of
RL Escherichia coli required for cell-cell interaction.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregory J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alha H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 12.7-28.0 min region on the linkage map.";
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregory J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsuo E., Nakayama K., Murata T., Tanaka M., Tode T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RL O157:H7 and genomic comparison with a laboratory strain K-12.";
RN [7]
RP SEQUENCE OF 22-346.
RC STRAIN-K12;
RX MEDLINE=81054820; PubMed=7001461;
RA Chen R., Schmidmayr W., Kramer C., Chen-Schmeissner U., Henning U.;
RT "Primary structure of major outer membrane protein II (ompA protein)
RL of Escherichia coli K-12.";
RN [8]
RP SEQUENCE OF 22-32.
RC STRAIN-K12 / W3110;
RA Pasquall C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.T.,
RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
RA Hochstrasser D.F.;
RT Submitted (SEP-1994) to the SWISS-PROT data bank.
RN [9]
RP SEQUENCE OF 22-34.
RC STRAIN-K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RL in the genome of Escherichia coli K-12.";
RN [10]
RP SEQUENCE OF 22-26.
RC STRAIN-K12 / W3110;
RX MEDLINE=98291876; PubMed=9629924;
RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traut M.,
RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;
RT "Extraction of membrane proteins by differential solubilization for
RL separation using two-dimensional gel electrophoresis.";
RN [11]
RP MUTANTS RESISTANT TO PHAGE ENTRY.
RX MEDLINE=84264337; PubMed=6086577;
RA Morona R., Klose M., Henning U.;
RT "Escherichia coli K-12 outer membrane protein (OmpA) as a
RL bacteriophage receptor: analysis of mutant genes expressing altered
RN [12]
RP J. Bacteriol. 159:570-578(1984).
RN [13]
RP PORIN ACTIVITY.
RC STRAIN-K12;
RX MEDLINE=92129334; PubMed=1370823;
RA Sugawara E., Nikaide H.;
RT "Pore-forming activity of OmpA protein of Escherichia coli.";
RL J. Biol. Chem. 267:2507-2511(1992).
RN [14]
RP TOPOLOGY.
RX MEDLINE=94148615; PubMed=8106193;
RA Gromiha M.M., Ponnuswamy P.K.;
RT "Prediction of transmembrane beta-strands from hydrophobic
RL characteristics of proteins.";
RN [15]
RP Int. J. Pept. Protein Res. 42:420-431(1993).
RN [15]
RP TOPOLOGY.
RX MEDLINE=99296577; PubMed=10368142;
RA Koebnik R.;
RT "Structural and functional roles of the surface-exposed loops of the
RL beta-barrel membrane protein OmpA from Escherichia coli.";
J. Bacteriol. 181:3688-3694(1999).

RL Nature 393:537-544(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z73101; CAA97374.1; -
 DR EMBL: AE006979; AAK45169.1; -
 DR TIGR: MT0922; -
 DR Tuberculist; Rv0899; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; FALSE_NEG.
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 30 50 POTENTIAL.
 FT TRANSMEM 101 121
 SQ SEQUENCE 326 AA; 33574 MW; 9E0D46ABCC179F4A CRC64;
 Query Match 16.6%; Score 143; DB 1; Length 326;
 Best Local Similarity 23.3%; Pred. No. 9.5e-06;
 Matches 52; Conservative 26; Mismatches 85; Indels 60; Gaps 4;
 QY 8 AAAAAALSVLTFTGCAKSTSOVMAP-----NATPGTGTG-----V 44
 DB 101 AAKAALMTALNGLADGVNVIDQIHDPVRSIDFSSAPVFAVPIPDFGLKVERDTV 160
 QY 45 IYGVAPLVNDDETFR-ALASKLPSTL----- 69
 DB 161 TLGTAPSSSHKRAVRAATSTWPMKIVNNIEVTCGAPPGPPASPCADLOSAINAYTG 220
 QY 70 --VYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDERGSRXYNMSLGERRAVA 127
 DB 221 GPITAFGNDGASLIPADYEILNRVADKLKACPDARVTINGTDTWTGSEGINIPLSAQRAKI 280
 QY 128 VRRVLLGKGINQASVELISFGERRPIAFGINNEAMSONRAEL 170
 DB 281 VADYLVARGVAGDHIAITVGISVNPPLASNATPEGRAKNRREVI 323
 RESULT 20
 OMPA_CITR STANDARD; PRT; 238 AA.
 AC P24016;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane protein A (Outer membrane protein II) (Fragment).
 GN OMPA.
 OS Citrobacter freundii.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Citrobacter.
 OC NCBI_Taxid=546;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OS60;

RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria.";
 RL J. Gen. Microbiol. 137:1911-1921(1991)
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF WAITING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTIONS (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M63354; AAA23095.1; -
 DR HSSP: P02934; 10JP
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OmpA_tmem.
 DR Pfam: PF00691; OmpA; 1.
 DR Pfam: PF01389; OmpA_membrane; 1.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; 1.
 DR Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
 KW NON_TER 1 1
 FT TRANSMEM <1 8 POTENTIAL.
 FT TRANSMEM 12 27 POTENTIAL.
 FT TRANSMEM 43 54 POTENTIAL.
 FT TRANSMEM 60 76 POTENTIAL.
 FT TRANSMEM 82 93 POTENTIAL.
 FT DOMAIN 104 109 3 X 2 AA TANDEN REPEATS OF A-P.
 FT REPEAT 104 105 1.
 FT REPEAT 106 107 2.
 FT REPEAT 108 109 3.
 FT DOMAIN 158 202
 FT DISULFD 212 224 OMPA-LIKE.
 FT NON_TER 238 238 BY SIMILARITY.
 SQ SEQUENCE 238 AA; 25664 MW; 66EA136D73A3F286 CRC64;
 Query Match 16.4%; Score 141; DB 1; Length 238;
 Best Local Similarity 28.3%; Pred. No. 9.9e-06;
 Matches 43; Conservative 24; Mismatches 61; Indels 24; Gaps 4;
 QY 43 GVIYTG-----APLVNDETFRALASK--LPSLVYFDPDSDEIKPQAAAILDE 89
 DB 82 GLISGVSVRFQGEAPVAVPAPAEVQTKHFTKSDVLFNFNKATLKEPQOALDQ 141
 QY 90 QAQFLTT--NOTARVLVAGHTDERGSRXYNMSLGERRAVNRNLLGKGINQASVELISF 147
 DB 142 MYSQLSNLDPKDGSVVVLFETDRIGSDAYNGLSEKRSYVVDYLISPSDKISARGM 201
 QY 148 GEERPIAFGTNE-----EAMSONRAEL 170
 DB 202 GESNPTVGTCDNVKARAALLDCLAPDRRVEI 233
 RESULT 21
 OMPA_SALTY STANDARD; PRT; 350 AA.
 AC P02936;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Outer membrane protein A precursor (Outer membrane protein 33K) (Outer
 DE membrane major heat-modifiable protein).
 GN OMPA OR STM1070.
 OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83287368; PubMed=6349993;
RA Freudl R., Cole S.T.;
RT "Cloning and molecular characterization of the ompa gene from
RT Salmonella typhimurium";
RL Eur. J. Biochem. 134:497-502(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lattelle P.,
RA Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2";
RL Nature 413:852-856(2001).
CC -I- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
CC SOLUTES (BY SIMILARITY).
CC -I- SUBUNIT: MONOMER (PROBABLE).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -I- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02006; CAA26037.1; -
DR EMBL; AE008746; AAL20003.1; -
DR PIR; A03436; MMEBAT.
DR HSSP; P02934; 1QUP.
DR Styene; SG10263; OMPA.
DR InterPro; IPR001145; Bac_OmpA.
DR InterPro; IPR000498; OmpA_tmam.
DR Pfam; PF01389; OmpA_membrane; 1.
DR ProDom; PD000930; Bac_OmpA; 1.
DR ProSite; PS01068; OMPA; 1.
KW Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
KW Porin; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 1 350
FT TRANSMEM 22 40
FT TRANSMEM 55 67
FT TRANSMEM 70 85
FT TRANSMEM 97 107
FT TRANSMEM 111 126
FT TRANSMEM 146 157
FT TRANSMEM 163 179
FT TRANSMEM 185 196
FT DOMAIN 205 212
FT REPEAT 205 212
FT REPEAT 205 206
FT REPEAT 207 208
FT REPEAT 209 210
FT REPEAT 211 212
FT DOMAIN 261 305
FT DISULFID 315 327
FT CONFLICT 114 114
FT CONFLICT 247 247
SQ SEQUENCE 350 AA; 37515 MW; BAAC52C8C5DE54FE CMC64;

Query Match 16.3%; Score 140.5; DB 1; Length 350;
Best Local Similarity 30.1%; Pred. No. 1.7e-05;
Matches 40; Conservative 20; Mismatches 62; Indels 11; Gaps 2;
OY 49 VAPLVNDNEFWKALSKPLSVYFDPDSDEKPKPQAAIILDRQAFLTT--NQTARVYVAG 106
DB 204 VAPAPAPPEVQVQTHFTLSKYVLEFNKSTLTPGQALQDLYLSQSLNLPKDGSVVVG 263
OY 107 HTDERGSKREYNMSLGERRAVAVRNYLLGKINGASVEIISFGEERPIAFGTNE----- 159
DB 264 FTRIGSDAVYNGLSERKAQSVVDYLISKIPSPKISARGGSENPYTGWTCOVKPRAA 323
OY 160 --EAMSONRRDEL 170
DB 324 LIDLIADPRVEI 336
RESULT 22
OMP_A_ESCHE
ID OMP_A_ESCHE STANDARD: PRT: 243 AA.
AC P24734;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Outer membrane protein A (Outer membrane protein II) (Fragment).
GN OMPA.
OS Escherichia hermannii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
OX NCBI_TaxID=565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33650, and ATCC 33652;
RX MEDLINE=92065252; PubMed=1955870;
RA Lawrence J.G., Ochman H., Hartl D.L.;
RT "Molecular and evolutionary relationships among enteric bacteria";
RL J. Gen. Microbiol. 137:1911-1921(1991).
CC -I- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
CC SOLUTES (BY SIMILARITY).
CC -I- SUBUNIT: MONOMER (PROBABLE).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -I- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63346; AAA24234.1; -
DR EMBL; M63347; AAA24238.1; -
DR HSSP; P02934; 1BXW.
DR InterPro; IPR001145; Bac_OmpA.
DR InterPro; IPR000498; OmpA_tmam.
DR Pfam; PF00691; OmpA; 1.
DR Pfam; PF01389; OmpA_membrane; 1.
DR ProDom; PD000930; Bac_OmpA; 1.
DR ProSite; PS01068; OMPA; 1.
KW Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
FT NON_TER 1 1
FT TRANSMEM <1 8
FT TRANSMEM 12 27
FT TRANSMEM 47 58
FT TRANSMEM 64 80
FT TRANSMEM 86 97
FT DOMAIN 107 114
FT REPEAT 107 108
FT REPEAT 109 110
SQ SEQUENCE 109 AA; 110

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X00254; CAA25062.1; -
 DR PIR: S07222; S07222.
 DR HSSP: P02934; 10JP.
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OmpA_tmern.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OmpA; 1.
 KW Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
 FT SIGNAL 1 21
 FT CHAIN 22 350
 FT TRANSMEM 27 40
 FT TRANSMEM 59 71
 FT TRANSMEM 74 89
 FT TRANSMEM 101 111
 FT TRANSMEM 115 130
 FT TRANSMEM 146 157
 FT TRANSMEM 163 179
 FT TRANSMEM 185 196
 FT DOMAIN 205 212
 FT REPEAT 205 206
 FT REPEAT 207 208
 FT REPEAT 209 210
 FT REPEAT 211 212
 FT DOMAIN 261 305
 FT DISULFID 315 327
 SQ SEQUENCE 350 AA; 37575 MW; 6276C6F2F21065DA CRC64;

Query Match 16.0%; Score 137.5; DB 1; Length 350;
 Best Local Similarity 31.5%; Pred. No. 3.2e-05;
 Matches 39; Conservative 16; Mismatches 58; Indels 11; Gaps 2;

QY 36 MAPGTGYITVAVLVNDETGVKALSKLPSELYFDFDSDEIKPQAAALIDEOAFLT 95
 DB 200 NAFV-----VAPAPAPAEVTKFTFLKSDVLFENKATLKPEGOALDOLYTQLS 250
 QY 96 T--NOTARVAVAGHTDERSREYNMSIGERRAVAVRNTYLLGGINOASVEIISFEERPI 153
 DB 251 NMDPKGSAVVLGYIDRISGEYINOKLSKRAQSVVDILVANGIPANKISARGMSDPV 310
 QY 154 AFGT 157
 DB 311 TGNT 314

RESULT 25
 ID OMPA_KLEPN STANDARD; PRT; 344 AA.
 AC P24017; 069435;
 DT 01-MAR-1992 (Rel. 21; Created)
 DT 15-DEC-1998 (Rel. 37; Last sequence update)
 DT 30-MAY-2000 (Rel. 39; Last annotation update)
 DE Outer membrane protein A precursor (Outer membrane protein II).
 GN OMPA.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RV 308;
 RX MEDLINE=98192544; PubMed=9524233;
 RA Nguyen T.N., Samuelson P., Sterky F., Merle-Polite C., Robert A.,
 RA Baussant T., Haeuw J.F., Uhlen M., Binz H., Stahl S.;
 RT "Chromosomal sequencing using a PCR-based biotin-capture method
 RT allowed isolation of the complete gene for the outer membrane protein

RT A of Klebsiella pneumoniae.";
 RL Gene 210:93-101(1998).
 RN [2]
 RP SEQUENCE OF 93-335 FROM N.A.
 RC STRAIN=LD119;
 RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria.";
 RL J. Gen. Microbiol. 137:1911-1921(1991).
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVEN LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ000998; CAA04450.1; -
 DR EMBL: M63355; AAN25119.1; -
 DR HSSP: P02934; 10JP.
 DR InterPro: IPR001145; Bac_OmpA.
 DR InterPro: IPR000498; OmpA_tmern.
 DR Pfam: PF00691; OmpA; 1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OmpA; 1.
 KW Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
 FT SIGNAL 1 2
 FT CHAIN 15 28
 FT TRANSMEM 48 60
 FT TRANSMEM 63 78
 FT TRANSMEM 90 100
 FT TRANSMEM 104 119
 FT TRANSMEM 140 151
 FT TRANSMEM 157 173
 FT TRANSMEM 179 190
 FT DOMAIN 199 206
 FT REPEAT 199 200
 FT REPEAT 201 202
 FT REPEAT 203 204
 FT REPEAT 205 206
 FT DOMAIN 255 299
 FT DISULFID 309 321
 FT CONFLICT 335 335
 SQ SEQUENCE 344 AA; 37061 MW; AC88AAE3B7871B16 CRC64;

Query Match 15.9%; Score 136.5; DB 1; Length 344;
 Best Local Similarity 30.2%; Pred. No. 3.9e-05;
 Matches 35; Conservative 20; Mismatches 50; Indels 11; Gaps 2;

QY 66 LPSLVYFDFDSDEIKPQAAALIDEOAFLT--NOTARVAVAGHTDERSREYNMSIGER 123
 DB 215 LKSDVLFENFKATLKPEGOALDOLYTQLSNDMPKGSAAVVLGYIDRISGEYINOKLSK 274
 QY 124 RAVAVRNTYLLGGINOASVEIISFEERPIAFTNE-----EAMSONRAEL 170
 DB 275 RAQSVVDYLVANGIPAGKISARGMSNVTGNTCDNVKARAALIDCLAPDRIVEI 330

RESULT 26
 OMPA_BUCAL

ID OMPA_BUCAT STANDARD: PRT; 349 AA.
AC P57414;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ompa-like protein precursor.
GN OMPA OR B0332.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: ACTS AS A PORIN WITH LOW PERMEABILITY THAT ALLOWS SLOW
CC PENETRATION OF SMALL SOLUTES (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (PROBABLY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AP001119; BAB3037.1; -
DR HSSP: P02934; 10JP.
DR InterPro: IPR001145; Bac_Ompa.
DR InterPro: IPR000498; Ompa_tmcm.
DR Pfam: PF00691; Ompa; 1.
DR Pfam: PF01389; Ompa_membrane; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_Ompa; 1.
DR PROSITE: PS01068; OMPA; FALSE_NEG.
KW Transmembrane; Porin; Signal; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 349 OMPA-LIKE PROTEIN.
FT TRANSMEM 27 40 POTENTIAL.
FT TRANSMEM 63 75 POTENTIAL.
FT TRANSMEM 78 93 POTENTIAL.
FT TRANSMEM 106 116 POTENTIAL.
FT TRANSMEM 120 135 POTENTIAL.
FT TRANSMEM 154 165 POTENTIAL.
FT TRANSMEM 171 187 POTENTIAL.
FT TRANSMEM 193 204 POTENTIAL.
FT DOMAIN 207 218 HINGE LIKE.
FT DOMAIN 267 311 OMPA-LIKE.
FT DISULFID 321 333 BY SIMILARITY.
SQ SEQUENCE 349 AA; 39303 MW; CC14AB1BD590CF58 CRC64;
Query Match 15.7%; Score 135.5; DB 1; Length 349;
Best Local Similarity 31.2%; Pred. No. 4.9e-05;
Matches 34; Conservative 23; Mismatches 45; Indels 7; Gaps 3;

RESULT 27
MOTY_VIBAN

ID MOTY_VIBAN STANDARD: PRT; 293 AA.
AC Q9S3P9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sodium-type flagellar protein moty precursor (Polar flagellum motor
GN MOTY.
OS Vibrio anguillarum (Listonella anguillarum).
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.
OX NCBI_TaxID=55601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NB10 / Serotype O1;
RA Ormande P., Milton D.L.;
RT "The role of motility in adherence and invasion of a fish cell line by
RT Vibrio anguillarum";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS THE ROLE OF A STATOR IN THE SODIUM FLAGELLAR
CC MOTOR, STABILIZING THE FORCE-GENERATING UNIT THROUGH DIRECT
CC INTERACTION WITH THE CELL WALL. INVOLVED IN ADHERENCE AND INVASION
CC OF FISH CELL LINE.
CC -1- SUBCELLULAR LOCATION: Membrane-associated (potential).
CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF176946; AAD51752.1; -
DR InterPro: IPR001145; Bac_Ompa.
DR InterPro: IPR001035; Moty.
DR Pfam: PF00691; Ompa; 1.
DR PRINTS: PR01023; NAFLGMOTY.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_Ompa; 1.
DR PROSITE: PS01068; OMPA; FALSE_NEG.
KW Flagella; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 293 SODIUM-TYPE FLAGELLAR PROTEIN MOTY.
FT DOMAIN 220 264 OMPA-LIKE.
SQ SEQUENCE 293 AA; 33725 MW; 5618435B4940D3B CRC64;
Query Match 14.8%; Score 127.5; DB 1; Length 293;
Best Local Similarity 28.6%; Pred. No. 0.00021;
Matches 30; Conservative 27; Mismatches 47; Indels 1; Gaps 1;
Db 68 SLVYPPDSDEIKRQAALIDDEQAQFLTNQRA-RVLVAGHTDERSGREYNSLGERAV 126
Db 183 TLHYDRDNVQLKASQKRAQLADYIRNODIDLVLSTYDVSQSDLSERRAE 242
Qy 127 AVENYLLGKINOASVEIISFGGEPIJAGTNEASQNRREL 171
Db 243 VLADYFKSLGLEDRIQVQCYGKRRPIADNASIGDKRRVVIS 287
RESULT 28
MOTY_VIBPA
ID MOTY_VIBPA STANDARD: PRT; 293 AA.
AC P46233;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Sodium-type flagellar protein moty precursor.
GN MOTY.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]

```

CC      1-1 SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC      -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its contents is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X00618; CAA25254.1; -.
DR      PIR; S07298; S07298.
DR      HSSP; P02934; 10UP.
DR      InterPro; IPR001145; Bac_Ompa.
DR      InterPro; IPR000498; Ompa_tmhm.
DR      Pfam; PF00691; Ompa; 1.
DR      Pfam; PF01389; Ompa_membrane; 1.
DR      PRINTS; PRO1021; OMPADOMAIN.
DR      ProDom; PD000930; Bac_Ompa; 1.
DR      PROSITE; PS01068; OMPA; 1.
KW      Outer membrane; Transmembrane; Phage recognition; Signal; Repeat;
KW      Portin.
FT      SIGNAL. 1 21
FT      CHAIN 22 359 OUTER MEMBRANE PROTEIN A.
FT      TRANSMEM 27 40 POTENTIAL.
FT      TRANSMEM 62 74 POTENTIAL.
FT      TRANSMEM 77 92 POTENTIAL.
FT      TRANSMEM 104 114 POTENTIAL.
FT      TRANSMEM 118 133 POTENTIAL.
FT      TRANSMEM 154 165 POTENTIAL.
FT      TRANSMEM 171 187 POTENTIAL.
FT      TRANSMEM 193 204 POTENTIAL.
FT      DOMAIN 210 219 5 x 2 AA TANDEM REPEATS OF A-P.
FT      REPEAT 210 211 1.
FT      REPEAT 212 213 2.
FT      REPEAT 214 215 3.
FT      REPEAT 216 217 4.
FT      REPEAT 218 219 5.
FT      DOMAIN 268 312 OMPA-LIKE.
FT      DISUFID 322 336 BY SIMILARITY.
SQ      SEQUENCE 359 AA; 38426 MW; 13992A037C19758B CRC64;

Query Match 14.2%; Score 122; DB 1; Length 359;
Best local Similarity 30.9%; Pred. No. 0.00084;
Matches 34; Conservative 19; Mismatches 49; Indels 8; Gaps 2;

QY 50 APFLVNDENVKALSKLPISLYPPFDDDEIKPQAAALIDEGQAQFLFT--NOTAFVLVGH 107
DB 218 AFVEVEKRT-----LSDVLEFNFNKSTLKAEGQALDLYTLQSSMPKGGSVVLCY 271
QY 108 TDEGRSREYNMSLGERRAVAARNYLLGKGINQASVELTISGEERPIAFCT 157
DB 272 TPVAGSDQYNOKLSFORAOSVVDYLVASKGPSKISARCGEADAVTGNT 321

RESULT 30
TP50_TREPA STANDARD; PRT; 417 AA.
ID TP50_TREPA STANDARD; PRT; 417 AA.
AC P38369; Q56358;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein TP50 precursor (Antigen TP57).
GN TP50 OR TP0292.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxId=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
MEDLINE=94156449; PubMed=8112835;
HA Hardham J.M.; Stamm L.V.;
NA

```

RT "Identification and characterization of the *Treponema pallidum* tps50 gene, an ompa homolog." Infect. Immun. 62:1015-1025(1994).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN-Nichols.

RC MEDLINE=95349399; PubMed=7623668;

RA Cox D.L., Atkins D.R., Porcella S.F., Norgard M.V., Radolf J.D.;

RT "Treponema pallidum in gel microdroplets: a novel strategy for investigation of treponemal molecular architecture." Mol. Microbiol. 15:1151-1164(1995).

RL [3]

RP SEQUENCE FROM N.A.

RC STRAIN-Nichols.

RA MEDLINE=98332770; PubMed=965876;

RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,

RA Dodson R., Gwin N., Hickey E.K., Clayton R., Ketchum K.A.,

RA Skolagen E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khatak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,

RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

RA Venter J.C.;

RT "Complete genome sequence of *Treponema pallidum*, the syphilis spirochete." Science 281:375-388(1998).

RL [1]

CC -1- FUNCTION: COULD FUNCTION AS A PORIN.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane (Probable).

CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U02628; AAA0517.2; -

DR EMBL: L28427; AAA99311.1; -

DR EMBL: AE001209; AAC65272.1; -

DR TIGR: TP0292;

DR InterPro: IPR001145; Bac_Ompa.

DR Pfam: PF00691; Ompa; 1.

DR ProDom: PD000930; Bac_Ompa; 1.

DR PROSITE: PS01068; OMPA; 1.

KW Transmembrane; Porlin; Signal; Outer membrane; Antigen;

KW Complete proteome.

FT STGM 1 26 POTENTIAL.

FT CHAIN 27 417 OUTER MEMBRANE PROTEIN TPNSO.

FT DOMAIN 345 390 OMPA-LIKE.

FT CONFLICT 127 127 MISSING (IN REF. 2).

FT CONFLICT 350 350 A -> E (IN REF. 2).

SO SEQUENCE 417 AA; 47608 MW; 079F579CFDD7288C CRC64;

Query Match 13.6%; Score 117; DB 1; Length 417;

Best Local Similarity 27.0%; Pred. No. 0.0029;

Matches 44; Conservative 22; Mismatches 65; Indels 32; Gaps 4;

QY 39 TGYTGYTYGVA-----PLYVDENVKALASLP-- 69

DB 253 SGGYLRKQGTATKANPFRPDPARVYQETLKLD--HMPDAKREEGVTISTEN 310

QY 70 YFDPDSDEIRKQAAAILDEQAFLETTNOTARVAVAGTGBERGSEKRENSLGRRAVAVR 129

DB 311 VQFDGASASLAPSEYEKRLTAIELLRFPPELLVSHAAARGSVQOQRISEBRADYVA 370

QY 130 NYLKGK--INQASVEIISFGEPRIATGNEANSQRRRAELS 171

DB 371 RYIQELGVADAHAHYTRCGAQOISIAFNDSSEDRKRRVREIT 413

RESULT 31

MOTB_BACSU

ID MOTB_BACSU STANDARD: PRT; 261 AA.

AC P28612;

DT 01-DEC-1992 (Rel. 24, Created)

DT 15-DEC-1992 (Rel. 24, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Chemotaxis motB protein (Motility protein B).

GN MOTB.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_Taxid=1423;

RP [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=92325000; PubMed=1624413;

RA Mirel D.B., Lustre V.M., Chamberlin M.J.;

RT "An operon of *Bacillus subtilis* motility genes transcribed by the sigma D form of RNA polymerase." J. Bacteriol. 174:4197-4204(1992).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN-168;

RA MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azavedo V., Berrero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Brouillet S., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,

RA Giuseppe G., Guy B.J., Hage K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Ilaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Kraerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestel D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogilvie A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetlelle D., Porwollik S., Prescott A.M.,

RA Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,

RA Sato T., Scanlan E., Schleich S., Schroeder R., Scifone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpiltra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,

RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*." Nature 390:249-256(1997).

RL Nature 390:249-256(1997).

CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE CELL WALL.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein.

CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M77238; AAA22603.1; -

DR EMBL: Z99111; CAB13241.1; -

DR PIR: S27516; S27516.

DR PIR: B42882; B42882.

DR Subtilist; BG10689; motB.

DR InterPro: IPR001145; Bac_Ompa.

DR Pfam: PF00691; OmpA: 1.
 DR ProDom: PD000930; Bac_OmpA: 1.
 KW Chemotaxis; Flagella; Transmembrane; Flagellar rotation;
 FT Complete proteome.
 FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 20 41 POTENTIAL.
 FT DOMAIN 261 261 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 261 261 BED623BE2A84CD5 CRC64;
 SQ SEQUENCE 261 AA; 29482 MW; 25972 MD; 687CC9C10DDEC64C CRC64;
 Query Match 12.4%; Score 106.5; DB 1; Length 261;
 Best Local Similarity 20.6%; Pred. No. 0.014; 79; Indels 71; Gaps 7;
 Matches 47; Conservative 31; Mismatches 79; Indels 71; Gaps 7;
 DB 1 MMLHIOIAAAALVLTFTMTGCAKSTQVAVAPNAPGTYGTGVI-YTGVA----- 51
 34 IYLVASSIDAARQML-----SKSFNEVF-----TGGTGLDYSSVTPPENESDGI 80
 52 -----LVNDN-----TYKALASKLP 67
 DB 81 DEVKKEKEKEKEKEKADAELEENKVSQVEKTKDKLEHOLETKMTSGLLITIK 140
 OY 68 SLVYFDEPDSDEIKPOAAALIDEOQFLTTNOTARVIVAGHTD-----ERGSREYNMGLGER 123
 DB 141 DSIFDSSKATIRKEDVPLAKEISNLVINPRTIIISGHTDMPKINSEFQSNMHLSTM 200
 OY 124 RAVAVRNYLIGKGINQASV-EIISGEERPIAGCTNEANSQNRRAEL 170
 DB 201 RAVNFMGLLIENPKLDAKVFSAKGYGKPYASNKTAEGRSKNRREV 248
 RESULT 32
 OMPA_SEROD STANDARD: PRT; 243 AA.
 ID OMPA_SEROD STANDARD: PRT; 243 AA.
 AC P24755;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Outer membrane protein A (outer membrane protein II) (Fragment).
 GN OMPA.
 OS Serratia odorifera.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 OX NCBI_TaxID=618;
 RN 111
 RP SEQUENCE FROM N.A.
 RX STRAIN-ATCC 3307;
 RX MEDLINE=92065252; PubMed=1955870;
 RA Lawrence J.G., Ochman H., Hartl D.L.;
 RT "Molecular and evolutionary relationships among enteric bacteria.";
 RT J. Gen. Microbiol. 137:1911-1921(1991).
 CC -1- FUNCTION: REQUIRED FOR THE ACTION OF COLICINS K AND L AND FOR THE
 CC STABILIZATION OF MATING AGGREGATES IN CONJUGATION. SERVES AS A
 CC RECEPTOR FOR A NUMBER OF T-EVER LIKE PHAGES. ALSO ACTS AS A PORIN
 CC WITH LOW PERMEABILITY THAT ALLOWS SLOW PENETRATION OF SMALL
 CC SOLUTES (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (PROBABLY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE OMPA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: M63357; AAA26561.1; -
 CC HSSP: P02934; 1QJP.
 CC InterPro: IPR001145; Bac_OmpA.
 CC InterPro: IPR000498; OmpA_tmam.
 DR Pfam: PF00691; OmpA: 1.
 DR Pfam: PF01389; OmpA_membrane; 1.

DR ProDom: PD000930; Bac_OmpA: 1.
 DR PROSITE: PS01068; OMPA; FALSE_NEG.
 KW Outer membrane; Transmembrane; Phage recognition; Repeat; Porin.
 FT NON_TER 1 1
 FT TRANSMEM 12 8 POTENTIAL.
 FT TRANSMEM 48 27 POTENTIAL.
 FT TRANSMEM 48 59 POTENTIAL.
 FT TRANSMEM 65 81 POTENTIAL.
 FT TRANSMEM 87 98 POTENTIAL.
 FT DOMAIN 104 113 5 X 2 AA TANDEM REPEATS OF A-P.
 FT REPEAT 104 105 1.
 FT REPEAT 106 107 2.
 FT REPEAT 108 109 3.
 FT REPEAT 110 111 4.
 FT REPEAT 112 113 5.
 FT DISULFID 215 229 BY SIMILARITY.
 FT NON_TER 243 243
 SQ SEQUENCE 243 AA; 25972 MW; 687CC9C10DDEC64C CRC64;
 Query Match 12.1%; Score 104; DB 1; Length 243;
 Best Local Similarity 29.1%; Pred. No. 0.022;
 Matches 34; Conservative 16; Mismatches 55; Indels 12; Gaps 2;
 DB 66 LPSLVFDEPDSDEIKPOAAALIDEOQFLTT--NOTARVIVAGHTDERSREYNMGLGER 123
 DB 122 LKSDVLEFNSAKSSLKPEGQALDLYTQSSMDPKGVSIVVAGTYDPVKGDAANGLSEA 181
 OY 124 RAVAVRNYLIGKGINQASV-EIISGEERPIAGCTNEANSQNRRAEL 170
 DB 182 RASVVDIVYSKGIIPAKIKISARGGEMADYDTSCGYKNGATFAQIECLAPRRREVI 238
 RESULT 33
 YTXE_BACME STANDARD: PRT; 218 AA.
 ID YTXE_BACME STANDARD: PRT; 218 AA.
 AC P46827;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 24.6 kDa protein in CCPA 3 region (ORF2).
 GN YTXE.
 OS Bacillus megaterium.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1404;
 RN 111
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94259294; PubMed=8200532;
 RA Hueck C., Kraus A., Hillen W.;
 RT "Sequences of ccpA and two downstream Bacillus megaterium genes with
 RT homology to the mcpA operon from Bacillus subtilis.";
 RT Gene 143:147-148(1994).
 CC -1- FUNCTION: MAY BE INVOLVED IN SOME TRANSPORT FUNCTION.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: L26052; AAA22297.1; -
 CC InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA: 1.
 DR ProDom: PD000930; Bac_OmpA: 1.
 KW Hypothetical protein; Transport; Transmembrane.
 FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 15 35 POTENTIAL.
 FT DOMAIN 36 218 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 36 218
 SQ SEQUENCE 218 AA; 24642 MW; 63622D730A4AE247 CRC64;

Query Match 11.6%; Score 100; DB 1; Length 218;
 Best Local Similarity 27.3%; Pred. No. 0.045;
 Matches 30; Conservative 17; Mismatches 57; Indels 6; Gaps 3;

CC
 CC EMBL: X01818; CA25961.1; ALT_INIT.
 CC DR EMBL: AE000346; AAC75654.1; -
 CC DR EMBL: D90888; BAA16490.1; -
 CC DR PIR: S07952; S07952.
 CC DR Ecogen: EG1152; yf1b.
 CC DR InterPro: IPR00145; Bac_OmpA.
 CC DR Pfam: PF00691; OmpA; 1.
 CC DR PRINTS: PR01021; OMPADOMAIN.
 CC DR ProDom: PD000930; Bac_OmpA; 1.
 CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; UNKNOWN.1.
 CC KW Membrane; Lipoprotein; Signal; Complete proteome.
 CC FT SIGNAL 18
 CC FT CHAIN 19 160
 CC FT LIPID 19 19
 CC FT POTENTIAL.
 CC FT PUTATIVE LIPOPROTEIN Yf1b.
 CC FT N-ACYL DIGLYCERIDE (POTENTIAL).
 CC SQ SEQUENCE 160 AA; 17244 MW; 33FE5FB174B7889 CRC64;

Query Match 11.4%; Score 98.5; DB 1; Length 160;
 Best Local Similarity 23.8%; Pred. No. 0.043;
 Matches 36; Conservative 25; Mismatches 65; Indels 25; Gaps 4;

CC
 CC EMBL: X01818; CA25961.1; ALT_INIT.
 CC DR EMBL: AE000346; AAC75654.1; -
 CC DR EMBL: D90888; BAA16490.1; -
 CC DR PIR: S07952; S07952.
 CC DR Ecogen: EG1152; yf1b.
 CC DR InterPro: IPR00145; Bac_OmpA.
 CC DR Pfam: PF00691; OmpA; 1.
 CC DR PRINTS: PR01021; OMPADOMAIN.
 CC DR ProDom: PD000930; Bac_OmpA; 1.
 CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; UNKNOWN.1.
 CC KW Membrane; Lipoprotein; Signal; Complete proteome.
 CC FT SIGNAL 18
 CC FT CHAIN 19 160
 CC FT LIPID 19 19
 CC FT POTENTIAL.
 CC FT PUTATIVE LIPOPROTEIN Yf1b.
 CC FT N-ACYL DIGLYCERIDE (POTENTIAL).
 CC SQ SEQUENCE 160 AA; 17244 MW; 33FE5FB174B7889 CRC64;

RESULT 34
 Yf1b_ECOLI
 ID Yf1b_ECOLI STANDARD; PRT; 160 AA.
 AC P07021;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative lipoprotein yf1b precursor.
 GN Yf1b OR B2605.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia
 CC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94057772; PubMed=6357787;
 RA Bystrom A.S., Hjalmarsson K.J., Wikstrom P.M., Bjork G.R.;
 RT "The nucleotide sequence of an Escherichia coli operon containing
 RT genes for the trnA(mis)methyltransferase, the ribosomal proteins S16
 RT and L19 and a 21-K polypeptide.";
 RL EMBO J. 2:899-905(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12 / MG1655;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1243-1248(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12;
 RA MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiuchi T.;
 RT "Construction of a contiguous 874-Kb sequence of the Escherichia coli
 RT K-12 genome corresponding to 50.0-66.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [4]
 RP SHOWS THAT THE PROTEIN IS EXPRESSED.
 RX MEDLINE=83141045; PubMed=6298574;
 RA Bystrom A.S., Bjork G.R.;
 RT "The structural gene (trnM) for the trnA(mis)methyltransferase is
 RT part of a four polypeptide operon in Escherichia coli K-12.";
 RL Mol. Genet. 188:447-454(1982).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC
 CC EMBL: X01818; CA25961.1; ALT_INIT.
 CC DR EMBL: AE000346; AAC75654.1; -
 CC DR EMBL: D90888; BAA16490.1; -
 CC DR PIR: S07952; S07952.
 CC DR Ecogen: EG1152; yf1b.
 CC DR InterPro: IPR00145; Bac_OmpA.
 CC DR Pfam: PF00691; OmpA; 1.
 CC DR PRINTS: PR01021; OMPADOMAIN.
 CC DR ProDom: PD000930; Bac_OmpA; 1.
 CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; UNKNOWN.1.
 CC KW Membrane; Lipoprotein; Signal; Complete proteome.
 CC FT SIGNAL 18
 CC FT CHAIN 19 160
 CC FT LIPID 19 19
 CC FT POTENTIAL.
 CC FT PUTATIVE LIPOPROTEIN Yf1b.
 CC FT N-ACYL DIGLYCERIDE (POTENTIAL).
 CC SQ SEQUENCE 160 AA; 17244 MW; 33FE5FB174B7889 CRC64;

Query Match 11.4%; Score 98.5; DB 1; Length 160;
 Best Local Similarity 23.8%; Pred. No. 0.043;
 Matches 36; Conservative 25; Mismatches 65; Indels 25; Gaps 4;

CC
 CC EMBL: X01818; CA25961.1; ALT_INIT.
 CC DR EMBL: AE000346; AAC75654.1; -
 CC DR EMBL: D90888; BAA16490.1; -
 CC DR PIR: S07952; S07952.
 CC DR Ecogen: EG1152; yf1b.
 CC DR InterPro: IPR00145; Bac_OmpA.
 CC DR Pfam: PF00691; OmpA; 1.
 CC DR PRINTS: PR01021; OMPADOMAIN.
 CC DR ProDom: PD000930; Bac_OmpA; 1.
 CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; UNKNOWN.1.
 CC KW Membrane; Lipoprotein; Signal; Complete proteome.
 CC FT SIGNAL 18
 CC FT CHAIN 19 160
 CC FT LIPID 19 19
 CC FT POTENTIAL.
 CC FT PUTATIVE LIPOPROTEIN Yf1b.
 CC FT N-ACYL DIGLYCERIDE (POTENTIAL).
 CC SQ SEQUENCE 160 AA; 17244 MW; 33FE5FB174B7889 CRC64;

RESULT 35
 MOTB_TREPA
 ID MOTB_TREPA STANDARD; PRT; 238 AA.
 AC C07887;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chemotaxis motB protein (Motility protein B).
 GN MOTB OR TP0724.
 OS Treponema pallidum.
 CC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 CC NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96345631; PubMed=8755894;
 RA Limberger R.J., Sliemers L.L., El-Afandi M.C.T., Dantuono L.A.;
 RT "Organization, transcription, and expression of the 5' region of the
 RT fla operon of Treponema phagedenis and Treponema pallidum.";
 RL J. Bacteriol. 178:4628-4634(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Nichols;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete.";
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT
 CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE
 CC CELL WALL (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.

CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U28219; AAB61254.1; -

DR TIGR: TP0724; AAC65689.1; -

DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA; 1.

DR ProDom: PD000930; Bac_OmpA; 1.

KW Chemotaxis; Flagella; Transmembrane; Inner membrane;

KW Flagellar rotation; Complete proteome.

FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 15 34 POTENTIAL.

FT DOMAIN 35 238 PERIPLASMIC (POTENTIAL).

SQ SEQUENCE 238 AA; 26050 MW; 0AB5FCDBE9852E1F CRC64;

Query Match 11.1%; Score 95.5; DB 1; Length 238;

Best Local Similarity 29.4%; Pred. No. 0.13;

Matches 37; Conservative 20; Mismatches 58; Indels 11; Gaps 5;

QY 53 VDNDETVKALASKLPDLVYFDFDSEIK-PQAAAILDQAOPLTTNQA--RVLYAGHTD 109

Db 108 YVSD--RGLVYISLTSDFYFGSSDLNVEESREALLRVQPLSHALAGRFRFEGHTD 165

QY 110 -----ERGSREYNMSIGERRAVAVRNLLGKINGOASVEIISFGERRIAGTNEAMSQ 164

Db 166 SVEVPDGGSTD-NMELSTRRAVRVLYHLYLDFGAQNRSLAGYADTRAFKFSNESSEGRAY 224

QY 165 NRRAPL 170

Db 225 NRRVVD 230

RESULT 36

MOTB_HELPY STANDARD: PRT; 257 AA.

ID MOTB_HELPY

AC P56427;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Chemotaxis motB protein (Motility protein B).

GN MOTB OR HP0816.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=26695 / ATCC 700392;

RX MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kervilave A.R., Clayton R.A., Sutton G.G.,

RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,

RA Loftus K., Richardson D., Dodson R., Khalak H.G., Glodek A.,

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Wathey L., Wallin E.,

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

RA Venter J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter

RT pylori.";

RL Nature 388:539-547(1997).

CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT

CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE

CC CELL WALL (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.

CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AE000593; AAD07865.1; -

DR TIGR: HP0816; -

DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA; 1.

DR ProDom: PD000930; Bac_OmpA; 1.

KW Chemotaxis; Flagella; Transmembrane; Inner membrane;

KW Flagellar rotation; Complete proteome.

FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 17 37 POTENTIAL.

FT DOMAIN 38 257 PERIPLASMIC (POTENTIAL).

SQ SEQUENCE 257 AA; 28849 MW; 58C1774B0547C032 CRC64;

Query Match 10.1%; Score 87; DB 1; Length 257;

Best Local Similarity 20.9%; Pred. No. 0.82;

Matches 43; Conservative 27; Mismatches 88; Indels 48; Gaps 6;

QY 13 ALSVTFMTGCAKSTSQ-----VWAPNAPPTGYGYITGYA 50

Db 29 ALFTALYLAISAVNKSVKALTEFTIKFNYPAPKPEAMQPVVYIPDGSKEEDOMSESSK 88

QY 51 PLVDNDETVKALAS-----KLPD-LVYDFDPSDEIKPQAAAILDQAOPL 93

Db 89 PASQNETETKATIKARKGVSYLEQIDGSLIKLSNLLFEMANSDAINODMMLYIEIKAKI 148

QY 94 L-TTNOTARVLVAGHTDE-----RGSREYNMSIGERRAVAVRNLLGKINGOASVEIIS 146

Db 149 IQKLPKRVHINVGFTDDPTLVKTRKSH--ELANRAVRVVKVLIQGVNPNQLSFSS 206

QY 147 FGEERPFAFTNEEAMSONRRAPLSY 172

Db 207 YGNTPIAPNDSLENRMKNRVEITF 232

RESULT 37

YTXE_BACSU STANDARD: PRT; 242 AA.

ID YTXE_BACSU

AC P39064;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein YTXE.

GN Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=95020526; PubMed=7934817;

RA Grundy F.J., Waters D.A., Takova T.Y., Henkin T.M.;

RA "Identification of genes involved in utilization of acetate and

RA acetoin in Bacillus subtilis.";

RL Mol. Microbiol. 10:259-271(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98048467; PubMed=9387221;

RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;

RA "Sequencing and functional annotation of the Bacillus subtilis genes

RA in the 200 kb rmb-dnaB region.";

RL Microbiology 143:3431-3441(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-168;
 RX MEDLINE-98044033; PubMed-9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcher S.,
 RA Boriss R., Bourcier L., Brans A., Braum M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaat A.,
 RA Hilbert H., Holst S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilawa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priesen E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivalet C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scifone F.,
 RA Sekiguchi J., Sekowska A., Serr S.J., Serr P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Vandenbol M., Vannier F., Vassart A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,
 RA Viari A., Wampat R., Wedler E., Wedler H., Weltzeneger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein E.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN SOME TRANSPORT FUNCTION.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).
 CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L17309; AAA68283.1;
 DR EMBL: AF008220; AAC00301.1;
 DR EMBL: 299119; CAB14950.1;
 DR PIR: S39642; S39642.
 DR Subtilisin; BG10366; YLX.
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR ProDom: PD000930; Bac_OmpA.1.
 KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
 FT DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 24 44 POTENTIAL.
 FT DOMAIN 45 242 EXTRACELLULAR (POTENTIAL).
 SQ SEQUENCE 242 AA; 27595 MW; 88A7DDC103CIDFE5 CRC64;

Query Match 10.0%; Score 86; DB 1; Length 242;
 Best Local Similarity 26.4%; Pred. No. 0.94; Indels 6; Gaps 3;
 Matches 29; Conservative 13; Mismatches 62;

OY 66 LPSLVYFDPDSDEIKPQAAALIDBQAFLLTNOTARVAVAGHTDERGSRVY---NNSIG 121
 DB 124 LQEAFLFTGTAKVKNLNETLLHQAVALQITPN-DIQVEGHTDSRNISTRTYSNNELS 182
 OY 122 ERRAVAVNYLGR-GINQASVEIISGEERPIAFGNEEAWSONRRRAEL 170
 DB 183 AARAGVIOYFTSKRKLPSKRFIAVGADTRKVPVNDKNTNEMKENRRVEI 232

RESULT 38

G6P2_STRCO
 ID G6P2_STRCO STANDARD; PRT; 551 AA.
 AC 095523;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Glucose-6-phosphate isomerase 2 (EC 5.3.1.9) (GPI 2) (phosphoglucose
 DE isomerase 2) (Pgi 2) (Phosphohexose isomerase 2) (Pgi 2).
 GN PGI2 OR SC01942 OR SC054.02C.
 OS Streptomyces coelicolor.
 CC Bacteria; Actinobacteria (class); Actinobacteridae;
 CC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_Taxid:1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RX MEDLINE-21996410; PubMed-12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrrell B.G., Parkhill J.,
 RT Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)".
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
 CC phosphate.
 CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AL035591; CAB38132.1;
 DR HSSP: 09N1E2; IHOX.
 DR InterPro: IPR01672; G6P_Isomerase.
 DR Pfam: PF00342; PGI.1.
 DR PRINTS: PR00662; G6PISOMERASE.
 DR PROSITE: PS00765; P_GLUCOSE_ISOMERASE_1; 1.
 DR PROSITE: PS00174; P_GLUCOSE_ISOMERASE_2; 1.
 KW Isomerase; Gluconeogenesis; Glycolysis; Complete proteome.
 FT ACT_SITE 390 390 BY SIMILARITY.
 FT ACT_SITE 514 514 BY SIMILARITY.
 SQ SEQUENCE 551 AA; 60569 MW; 631BD4A76EB34643 CRC64;

Query Match 9.6%; Score 83; DB 1; Length 551;
 Best Local Similarity 25.0%; Pred. No. 4.7;
 Matches 23; Conservative 14; Mismatches 53; Indels 2; Gaps 1;

OY 30 QVAVAPAPPGYGVYITGVAPLVNDDEFYKALASPLVYFDPDSDEIKPQAAALIDE 89
 DB 69 QELAAATGVSGGLDAMFGERIRITEDRAVLHTALRAPDAVLEGVENVQVAVADLK 128
 OY 90 QAOFLTNOTARVAVAGHTDERGSRVY---NNSIG 121
 DB 129 MAGFADRVKGE--WTGHTGRRIRVNVNIGIG 158

RESULT 39
 MOTB_HELPJ STANDARD; PRT; 257 AA.
 ID MOTB_HELPJ
 AC 092129;
 DT 16-OCT-2001 (Rel. 40, Created)

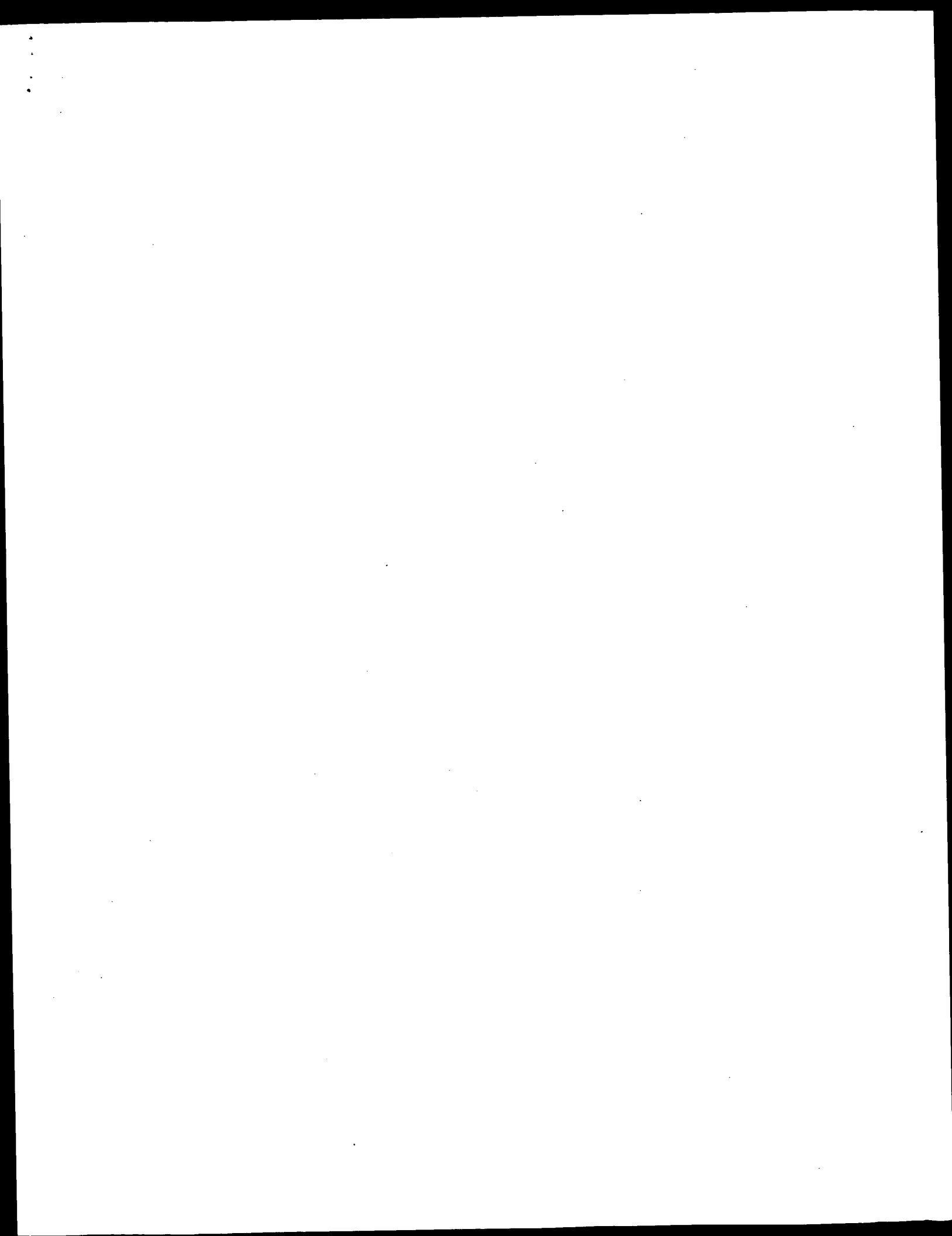
DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Chemotaxis motB protein (Motility protein B).
 GN MOTB OR JHP0752.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RX NCBI_TaxID=85963;
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tammino P.J., Caruso A., Uria-Nikelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 357:176-180(1999).
 CC -1- FUNCTION: REQUIRED FOR THE ROTATION OF THE FLAGELLAR MOTOR. MIGHT
 CC BE A LINKER THAT FASTENS THE TORQUE-GENERATING MACHINERY TO THE
 CC CELL WALL (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE MOTB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001506; AAC06331.1; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA: 1.
 DR ProDom: PD000930; Bac_OmpA: 1.
 KW Chemotaxis; Flagella; Transmembrane; Inner membrane;
 KW Flagellar rotation; Complete proteome.
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 37 POTENTIAL.
 FT DOMAIN 38 257 PERIPLASMIC (POTENTIAL).
 FT SEQUENCE 257 AA; 28838 MW; 04F8E7BF833A9FB CRC64;
 SQ
 Query Match 9.5%; Score 82; DB 1; Length 257;
 Best Local Similarity 20.4%; Pred. No. 2.3;
 Matches 42; Conservative 25; Mismatches 91; Indels 48; Gaps 5;
 QY 13 ALSVLTFMTCANKSTSQ-----VMAAPAPGYGYITGYA 50
 DB 29 ALFTALYSAVSKSVKAEKTEFIENYAPRPEAMOPVVIIPDSGKEBOMASESK 88
 QY 51 PLVNDDETVALASKLPSLVYFPDSDDEIKPQAAILDQAOFLTNQ-----TARV 102
 DB 89 PASQNETTKATIKRKGSLVLEQIDGSLVKLPSSLLFNANSDAIDNOMLIEIETIAT 148
 QY 103 L-----VAGHTDE-----KGSREYNMSLGERRAVAVRNYLLGKINGQASVEIIS 146
 DB 149 IQKLPRKHIVNGFTDNTPLNKTREKSHY--ELANRAVRYKVKVLIQGVDPNQISFSS 206
 QY 147 FGEERPIATCTNEEAMSONRAELSY 172
 DB 207 YGSTNPAPNDLSLENRRKNNRVEIFF 232
 RESULT 40
 COOS_RHORI STANDARD; PRT; 639 AA.
 AC P31896;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Carbon monoxide dehydrogenase (EC 1.2.99.2) (CODH).

GN COOS.
 OS Rhodospirillum rubrum.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 CC Rhodospirillum.
 RX NCBI_TaxID=1085;
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RA STRAIN=URL;
 RA MEDLINE=92355502; PubMed=1644755;
 RA Kerby R.L., Hong S.S., Ensign S.A., Coppoc L.J., Ludden P.W.,
 RA Roberts G.P.;
 RT "Genetic and physiological characterization of the Rhodospirillum
 RT rubrum carbon monoxide dehydrogenase system.";
 RL J. Bacteriol. 174:5284-5294(1992).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=URL;
 RX MEDLINE=96198151; PubMed=862676;
 RX Fox J.D., Kerby R.L., Roberts G.P., Ludden P.W.;
 RT "Characterization of the CO-induced, CO-tolerant hydrogenase from
 RT Rhodospirillum rubrum and the gene encoding the large subunit of the
 RT enzyme.";
 RL J. Bacteriol. 178:1515-1524(1996).
 CC -1- FUNCTION: CODH MAY OXYDATE CARBON MONOXIDE COUPLED, VIA COOF, TO
 CC REDUCTION OF HYDROGEN CATION BY AN HYDROGENASE (POSSIBLY COOH).
 CC -1- CATALYTIC ACTIVITY: CO + H(2)O + acceptor -> CO(2) + reduced
 CC acceptor.
 CC -1- COFACTOR: IRON, NICKEL AND ZINC.
 CC -1- INDUCTION: By carbon monoxide; under anaerobic conditions.
 CC -1- MISCELLANEOUS: METHYL VIOLOGEN CAN ACT AS ACCEPTOR.
 CC -1- SIMILARITY: STRONG, TO C.THERMILACETICUM CARBON MONOXIDE
 CC DEHYDROGENASE BETA SUBUNIT AND SOME, TO M.SOEHNENII CARBON
 CC MONOXIDE DEHYDROGENASE ALPHA SUBUNIT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U65510; AAC5123.1; -
 DR PIR: C42957; C42957.
 DR InterPro: IPR004137; Prismane.
 DR Pfam: PF03063; Prismane: 1.
 KW Oxidoreductase; Nickel; Iron; Zinc; Iron-sulfur.
 FT METAL 50 50 IRON-SULFUR (BY SIMILARITY).
 FT METAL 53 53 IRON-SULFUR (BY SIMILARITY).
 FT METAL 72 72 IRON-SULFUR (BY SIMILARITY).
 FT SEQUENCE 639 AA; 66854 MW; DCADDTC13D8D588 CRC64;
 SQ
 Query Match 9.4%; Score 81; DB 1; Length 639;
 Best Local Similarity 22.6%; Pred. No. 8.6;
 Matches 42; Conservative 26; Mismatches 58; Indels 60; Gaps 8;
 QY 4 HIQIAAAALSVLFMTGANKSTSOV--WAPAPPYGYGYITGY-----AP 51
 DB 462 YVDLAKSLAKRNVLVLATGCAAGAFAPAGALMTSEATQYAGELKGLVSAIGTAAGIGCP 521
 QY 52 L-----VDNDETVALASKLPSLVYFPDSDDEIKPQAAILDQAO----- 91
 DB 522 LPLVHMGSQVNDNRV--ALATVALANKIGVLDLPLVASAPCEMSKALAGSWAVTIG 580
 QY 92 -----QFLT--TNOTARVYLVAGH-----TDERGSGREYNMSLGERRAVAVRNY 131
 DB 581 LPTHGVSYPYIGSOIYTKVLTETAKDLVGSYFIVDIDPK-----SADKLYAAIQER 633
 QY 132 LLGKGI 137
 DB 634 RAGIGL 639

Mon Jul 7 08:55:32 2003

Search completed: July 6, 2003, 14:05:18
Job time : 39 secs

us-09-674-779b-2.rsp



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:04:35 / Search time 58 Seconds

(Without alignments)
611.036 Million cell updates/sec

Title: US-09-674-779B-2

Sequence: 1 MMHQTAAALSVTFM.....IAFGTEAMSONRAELSY 172

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Listing first 100 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	307	35.7	168	2	051489
2	306	35.5	168	16	091424
3	289	33.6	174	16	08XRH6
4	287.5	33.4	168	2	093455
5	286.5	33.3	172	16	081FE4
6	286.5	33.3	188	16	09A5H5
7	283	32.9	168	16	08ZG20
8	269.5	31.3	168	16	09BFR5
9	265.5	30.8	186	16	09PC85
10	262.5	30.5	176	16	09Z6C3
11	262	30.4	177	16	08U9L5
12	259	30.1	172	16	09KR12
13	258	30.0	130	2	09L308
14	256	29.7	155	16	09ZCH2
15	254	29.5	168	16	044662
16	253.5	29.4	152	2	044157

17	246	28.6	157	2	047958	047958 haemophilus
18	242.5	28.2	155	16	0928C3	0928C3 rickettsia
19	235	27.3	171	2	045017	045017 bartonella
20	215.5	25.0	179	16	025750	025750 helicobacte
21	215.5	25.0	179	16	092K87	092K87 helicobacte
22	208.5	24.2	179	2	092AW7	092AW7 helicobacte
23	206.5	24.0	209	16	08XWU0	08XWU0 rickettsia
24	205.5	23.9	157	2	09RNP4	09RNP4 campylobact
25	204.5	23.8	165	16	046123	046123 campylobact
26	204.5	23.8	192	16	0927C5	0927C5 campylobact
27	204	23.7	221	16	092ST9	092ST9 rhizobium m
28	202	23.5	216	16	096NA9	096NA9 rhizobium l
29	201.5	23.4	249	2	08VR19	08VR19 myxococcus
30	199.5	23.2	165	2	046099	046099 campylobact
31	196.5	22.8	188	16	084605	084605 chlamydia t
32	196.5	22.8	202	16	09P7P0	09P7P0 chlamydia m
33	193.5	22.5	321	2	09XCZ6	09XCZ6 pseudomonas
34	191.5	22.2	321	2	09XCZ3	09XCZ3 pseudomonas
35	191	22.2	323	2	09XC25	09XC25 pseudomonas
36	189.5	22.0	261	16	09HXU8	09HXU8 pseudomonas
37	189	22.0	220	16	08UBX6	08UBX6 agrobacteri
38	189	22.0	277	16	08XR50	08XR50 rickettsia s
39	187.5	21.8	232	16	08YHL1	08YHL1 brucella me
40	187.5	21.8	328	2	09XC27	09XC27 pseudomonas
41	184	21.4	325	2	09X4R8	09X4R8 pseudomonas
42	180.5	21.0	318	2	09X4S1	09X4S1 pseudomonas
43	180	20.9	237	16	0915A7	0915A7 pseudomonas
44	180	20.9	242	16	09PGD9	09PGD9 pseudomonas
45	180	20.9	317	2	051781	051781 pseudomonas
46	179.5	20.8	153	2	09Y691	09Y691 helicobacte
47	179	20.8	317	2	051780	051780 pseudomonas
48	179	20.8	317	2	051721	051721 pseudomonas
49	179	20.8	326	2	09S4R5	09S4R5 pseudomonas
50	179	20.8	389	16	09PGF9	09PGF9 xylella fas
51	176.5	20.5	220	2	09XAX8	09XAX8 pseudomonas
52	176	20.4	325	2	09X4R7	09X4R7 pseudomonas
53	175	20.3	321	2	09WMT5	09WMT5 pseudomonas
54	174.5	20.3	151	16	08R6T1	08R6T1 pseudomonas
55	174.5	20.3	157	16	08R6N8	08R6N8 fusobacteri
56	174.5	20.3	321	16	031154	031154 vibrio chol
57	174	20.2	317	2	051777	051777 pseudomonas
58	173	20.1	91	2	048265	048265 haemophilus
59	173	20.1	407	16	09AA59	09AA59 caulobacter
60	171.5	19.9	269	16	09HZU7	09HZU7 pseudomonas
61	171.5	19.9	338	2	09WMT9	09WMT9 pseudomonas
62	169.5	19.7	226	16	09ABK6	09ABK6 caulobacter
63	169.5	19.7	343	2	09S6B8	09S6B8 pseudomonas
64	169	19.6	315	16	08REH8	08REH8 fusobacteri
65	169	19.6	449	16	09A2R2	09A2R2 caulobacter
66	168	19.5	346	2	051841	051841 actinobacti
67	167.5	19.5	344	2	09X4S3	09X4S3 pseudomonas
68	167.5	19.5	346	2	09S5Y9	09S5Y9 actinobacti
69	167	19.4	172	16	08RG63	08RG63 pseudomonas
70	166.5	19.3	220	2	09SOK6	09SOK6 shewanella
71	166.5	19.3	353	16	09CMM1	09CMM1 pasteurella
72	165.5	19.3	360	2	086254	086254 haemophilus
73	165	19.3	346	2	09X4R9	09X4R9 pseudomonas
74	165	19.2	317	2	051778	051778 pseudomonas
75	165	19.2	317	2	052465	052465 pseudomonas
76	164.5	19.1	326	2	09X4S0	09X4S0 pseudomonas
77	164	19.0	319	16	09PMI4	09PMI4 campylobact
78	163	18.9	273	2	09K558	09K558 rhodospirill
79	162.5	18.9	322	2	09X4R6	09X4R6 pseudomonas
80	162	18.8	343	2	09XCZ4	09XCZ4 pseudomonas
81	161.5	18.8	344	2	08RMM1	08RMM1 haemophilus
82	161.5	18.8	344	2	08RMM1	08RMM1 haemophilus
83	161.5	18.8	344	2	08RMM1	08RMM1 haemophilus
84	161.5	18.8	344	2	08RMM1	08RMM1 haemophilus
85	161	18.7	152	2	09S3M1	09S3M1 bordetella
86	161	18.7	348	2	09ZID6	09ZID6 methylococc
87	160.5	18.6	348	2	08VPL3	08VPL3 pasteurella
88	160.5	18.6	403	2	096773	096773 haemophilus
89	160	18.6	202	16	08RE60	08RE60 fusobacteri

08xfe2 salmonella
006895 campylobact
09x4s2 pseudomonas
09x4l6 pseudomonas
08xjdl haemophilus
08xjdl haemophilus
09x1t2 vibrio chol
0914t3 pseudomonas
0914t3 pseudomonas
08x158 haemophilus
052584 pseudomonas

ALIGNMENTS

RESULT 1
ID 051489 PRELIMINARY; PRT; 168 AA.
AC 051489:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Outer membrane protein.
GN OPR1.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97312009; PubMed=9168620;
RA Lim A., De Vos D., Brauns M., Mossialos D., Gaballa A., Hamers R.,
RA Cornelis P.;
RT *Molecular and immunological characterization of OprL, the 18 kDa
RT outer-membrane peptidoglycan-associated lipoprotein (Pal) of
RT Pseudomonas aeruginosa.
RT Microbiology 143:1703-1716(1997).
RL EMBL; 250191; CAA90573.1; -;
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_OmpA; 1.
DR PROSITE: PS01068; OMPA; 1.
SQ SEQUENCE 168 AA; 17898 MW; 74831279BBAH1186 CRC64;
Query Match 35.7%; Score 307; DB 2; Length 168;
Best Local Similarity 40.9%; Pred. No. 7.4e-21;
Matches 70; Conservative 39; Mismatches 44; Indels 18; Gaps 6;
QY 6 QIAAALSVLTFTMGCAK-----STQVWVAPNAPGTGVTGYVAVPLVNDDET 59
DB 8 KFAALALAMAV---AVGSSKSGDASGEGANGVDPN--GY-GANSAGVDSLSDEAL 61
QY 60 KALASKLPILVYFEDSDSEIKPOAAALDEQAQFLTTNOTARVLVAGHTDGRSEYKMS 119
DB 62 RAITT-----FYFEDSDSLKPEAMRALDVHAKDL-KSGGQRYVLEGHTDGRREYNMA 115
QY 120 LGERRAVAVRNYLIGKGINQASVEIISFGGERIATGTEASQNRRAEL 170
DB 116 LGERRAKAVORYLVLOGVSPALTELVSIGKERPVATGHDOSMAQNRREL 166
RESULT 2
ID 091424 PRELIMINARY; PRT; 168 AA.
AC 091424:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Outer membrane protein oprL precursor.
GN OPR1 OR PA0973.
OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT *Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004530; AAG04362.1; -;
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac_OmpA; 1.
DR PROSITE: PS01068; OMPA; 1.
KW Complete proteome.
SQ SEQUENCE 168 AA; 17925 MW; 03F36279BBAH1183 CRC64;
Query Match 35.5%; Score 306; DB 16; Length 168;
Best Local Similarity 40.9%; Pred. No. 9.2e-21;
Matches 70; Conservative 38; Mismatches 45; Indels 18; Gaps 6;

QY 6 QIAAALSVLTFTMGCAK-----STQVWVAPNAPGTGVTGYVAVPLVNDDET 59
DB 8 KFAALALAMAV---AVGSSKSGDASGEGANGVDPN--GY-GANSAGVDSLSDEAL 61
QY 60 KALASKLPILVYFEDSDSEIKPOAAALDEQAQFLTTNOTARVLVAGHTDGRSEYKMS 119
DB 62 RAITT-----FYFEDSDSLKPEAMRALDVHAKDL-KSGGQRYVLEGHTDGRREYNMA 115
QY 120 LGERRAVAVRNYLIGKGINQASVEIISFGGERIATGTEASQNRRAEL 170
DB 116 LGERRAKAVORYLVLOGVSPALTELVSIGKERPVATGHDOSMAQNRREL 166
RESULT 3
QXFEH6
ID 08XFEH6 PRELIMINARY; PRT; 174 AA.
AC 08XFEH6:
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Tol protein required for outer membrane integrity, uptake of group A
DE colicins, and translocation of phage DNA to cytoplasm
DE (peptidoglycan-associated lipoprotein).
OS Salmonella typhi.
OS Salmonella typhimurium, and
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES-S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX SPECIES-S.typhi; STRAIN=CT18;

MEDLINE-21534947; PubMed-11677608;
 RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typh CT18.",
 RL Nature 413:848-852(2001).
 DR EMBL: AE008730; AAL29693.1;
 DR EMBL: AL627268; CAD05211.1;
 DR InterPro: IPR001145; Bac_Ompa.
 DR Pfam: PF00691; Ompa.1.
 DR PRINTS: PRO1021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_Ompa.1.
 DR PROSITE: PS01068; OMPA.1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 174 AA; 18865 MW; A515624F96EB5A5 CRC64;

Query Match 33.6%; Score 289; DB 16; Length 174;
 Best Local Similarity 36.8%; Pred. No. 3.7e-19;
 Matches 67; Conservative 33; Mismatches 64; Indels 18; Gaps 3;

QY 1 MMLHIQIAAAALSVLTFTMGCAKSTQVWVAPNAPGYTGVIYGVAPLVDD----- 54
 DB 1 MMLHIQIAAAALSVLTFTMGCAKSTQVWVAPNAPGYTGVIYGVAPLVDD----- 54
 QY 55 --NDETVKALASLP--SLVYFDFDSEIKPQAAALDEQAFLTTNOTARVLVAGHTDE 110
 DB 53 MSSEBOARLOMOQLQONNTVYFDLKYDIRSDFAAMLDHANFLRNSPKYVGHADDE 112
 QY 111 RGSREXNMSIGERRAVAVNYLKGKINGQASVEIISFGGERPIAGTNEAASQRRREL 170
 DB 113 RGPPEYNISLIGERRAVALVYLGKRGVSADQISYSGKEKPAVLGHDEAAVAKNRRAVL 172
 QY 171 SY 172
 DB 173 VY 174

RESULT 4
 0934G5 PRELIMINARY; PRT; 168 AA.
 AC 0934G5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Peptidoglycan-associated lipoprotein precursor.
 GN PAL.
 OS *Erwinia chrysanthemi*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Peptobacterium.
 OX NCBI_TaxID=556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3937;
 RA Ray M.C., Vianney A., Cotte-pattat N., Lazzaroni J.;
 RT "Characterization of the *Erwinia chrysanthemi* tol-pal genes."
 RL Submitted (Oct-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ297885; CAC82710.1;
 DR InterPro: IPR001145; Bac_Ompa.
 DR Pfam: PF00691; Ompa.1.
 DR ProDom: PD000930; Bac_Ompa.1.
 DR PROSITE: PS01068; OMPA; UNKNOWN_1.
 DR Lipoprotein; Signal.
 KW SIGNAL.
 FT CHAIN 1 21 POTENTIAL.
 SQ SEQUENCE 168 AA; 18433 MW; 1E315BCF52852D55 CRC64;

Query Match 33.4%; Score 287.5; DB 2; Length 168;

Best Local Similarity 37.4%; Pred. No. 4.9e-19;
 Matches 65; Conservative 31; Mismatches 61; Indels 17; Gaps 3;
 QY 1 MMLHIQIAAAALSVLTFTMGCAKSTQVWVAPNAPGYTGVIYGVAPLVDDDETVK 60
 DB 10 LMLDLPLVAVAAGS-----NKRANNDOSSLNGAGMEN-----GNNSSAEQAR 54
 QY 61 ALASKLP--SLVYFDFDSEIKPQAAALDEQAFLTTNOTARVLVAGHTDERSREYNM 118
 DB 55 LQWELQNNIVYFDLKYDIRDFQOMLDHANFLRNSPKYVGHADDEGTPPEYNI 114
 QY 119 SLGERRAVAVNYLKGKINGQASVEIISFGGERPIAGTNEAASQRRREL 172
 DB 115 ALGERRAVAVNYLKGKINGQASVEIISFGGERPIAGTNEAASQRRREL 168

RESULT 5
 08Y1F4 PRELIMINARY; PRT; 172 AA.
 AC 08Y1F4;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE Probable peptidoglycan-associated lipoprotein precursor.
 GN PAL OR RSC0736 OR RS05117.
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 CC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1000;
 RX MEDLINE-21681879; PubMed-11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Atilat M., Billault A., Brotier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choise N., Claudet-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*."
 RL Nature 415:497-502(2002).
 DR EMBL: AL646060; CAD14266.1;
 DR InterPro: IPR000104; Antifreeze_1.
 DR ProDom: PD000930; Bac_Ompa.1.
 DR PROSITE: PS01068; OMPA; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 172 AA; 18656 MW; 5CFE28E6AE7702D9 CRC64;

Query Match 33.3%; Score 286.5; DB 16; Length 172;
 Best Local Similarity 38.2%; Pred. No. 6.3e-19;
 Matches 65; Conservative 36; Mismatches 62; Indels 7; Gaps 3;

QY 5 IDIAAAALSVLTFTMGCAKSTQVWVAPNAPGYTGVIYGVAPLVDDDETVKAL 62
 DB 8 IKLAIAIALLDLGGACSSSVKLDITRSK-----NATGAAAGADPRNVPYVSDDELTDPN 62
 QY 63 ASKLP--SLVYFDFDSEIKPQAAALDEQAFLTTNOTARVLVAGHTDERSREYNM 122
 DB 63 SPLAKRSYFDFDSTYKPEYGGILLTHARLQSHNRKYLIGQNTDERSTETNIALQ 122
 QY 123 RRAVAVNYLKGKINGQASVEIISFGGERPIAGTNEAASQRRREL 172
 DB 123 KRAEVRRAISLGVDPDSQMSVSLGKEKPAVLGHDEAASQRRREL 172

RESULT 6
 09A3H5 PRELIMINARY; PRT; 188 AA.
 AC 09A3H5;
 ID 09A3H5;
 SQ SEQUENCE 188 AA; 18433 MW; 1E315BCF52852D55 CRC64;

Query	Match	32.9%	Score 283;	DB 16;	Length 168;	
Best Local Similarity	37.9%	Pred. No. 1.3e-18;				
Matches	66;	Conservative 32;	Mismatches 68;	Indels 8;	Gaps 2;	
Db	1 MMLHIDIAAAAAALSVLTFEMTGCANKSTSOYVAVAPNAPGTGYIYGVAPLVNDDETVK	60				
Oy	1 MMLHIDIAAAAAALSVLTFEMTGCANKSTSOYVAVAPNAPGTGYIYGVAPLVNDDETVK	60				
Db	1 MMLHIDIAAAAAALSVLTFEMTGCANKSTSOYVAVAPNAPGTGYIYGVAPLVNDDETVK	54				
Oy	61 ALASKLP--SLVYFDFDSDEIKPQAAAILDEQAOEFLTTNQIARVLVAGHIDEGSGREYNN	118				
Db	55 LQWQELQKNNIYVFGFDKDYIGSDPDAQOMLDAAALFRSNPDKVYVGHADGKSTPEYNI	114				
Oy	119 SLGERAAVAVRNILYLGKGINQASVELISNGEERPIAFGTNEASQNRRAELSY	172				
Db	115 ALGERRASAVKMYLQCKGVASADQISTVSGKEKEPAVLGHDEDAAPAKNRRAVLVY	168				
RESULT 8						
ID Q98F85	PRELIMINARY;	PRT; 168 AA.				
Q98F85						
Q98F85;						

Df
Df
Df
DE
GN
OS
OC
OX
RN
RP
RC
RX
RA
RA
RA
RA
RT
RT
RL
DR
DR
DR
DR
DR
KW
SQ

01-OCT-2001 (Tremblérel_18, last sequence update)
01-OCT-2001 (Tremblérel_16, last sequence update)
01-MAR-2002 (Tremblérel_20, last annotation update)

Hypothetical protein ml13887.
ml13887.
Rhizobium lotfi (Mesorhizobium lotfi).
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
NCBI_Taxid=381;
[1]
SEQUENCE FROM N.A.
STRAIN=MAFF303099;
MEDLINE=21082930; PubMed=11214968;
Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kiyojima C., Kohara M., Matsumoto M., Matsuno A.,
Mochizuchi Y., Nayakaya S., Nakazaki N., Shimpo S., Sugimoto M.,
Takuchuki C., Yamada M., Tabata S.;
"Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium lotfi.";
DNA Res. 7:331-338(2000).
EMBL: AP003003; BAB50682.1; -
InterPro: IPR01145; Bac_CmpA.
Pfam: PF00691; ompA; 1
PRINTS: PRO1021; OMPADOMAIN.
ProDom: PD00930; Bac_OmpA; 1.
Hypothetical protein; Complete proteome.
SEQUENCE 168 AA; 17765 MW; 03E260D3BA1ECBB3 CRC64;

Query Match 31.3%; Score 269.5; DB 16; Length 168;
Best Local Similarity 40.3%; Pred.No.2,je-17;
Matches 64; Conservative Matches 53; Indels 17; Gaps

QY 13 ALSVLTFTMGTCANNSTGSVMVAPN--APTGYGTYYITGVA-PLVNDDEVVKALASKLSTL 69
DB 15 ALVAMATAIAGCAASKKT-----PNNAADLGING---AGAFTPSQAODFIV-----NIGR 60
QY 70 VFEPDDSDSEIKRQAAAILIDEOAFLLTNQTARVLVAGHTDERGSRREYNLSGERAVAVR 12
DB 61 IFEDDSSIRDAOTTLTARAQWMNYKYAIAYVEGHADERETREYNATLGRRAAAR 12
QY 130 NYLTKGIQAOASVELISIPESRRPIAFTGNEAWMSQNRA 168

Db 121 DFLVSKGVASSRLTTSYKREKREAVACDDISCSQNRRA 159

RESULT 9

09PC85 PRELIMINARY; PRT; 186 AA.

AC 09PC85; 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Outer membrane protein p6 precursor.

OS XylE1896.

OC XylE1896.

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC XylE1896.

OC NCBI_TaxID=2371;

RP STRAIN=9A5C;

RC MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Aruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Pelxoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA da Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
 RA de Souza A.P., Terenzi M.L.Z., Siqueira W.J. de Souza A.A.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Melandris J., Setubal J.C.,
 RA "The genome sequence of the plant pathogen Xylella fastidiosa.",
 RT Nature 406:151-159(2000).
 RL EMBL: AE004009; AAF84702.1; -
 DR InterPro: IPR001145; Bac_Ompa.
 DR Pfam: PF00691; Ompa.1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_Ompa.1.
 DR Complete proteome.
 SQ SEQUENCE 186 AA; 20431 MW; CDDDF6233DB0C92 CRC64;

Query Match 30.8%; Score 265.5; DB 16; Length 186;
 Best Local Similarity 35.9%; Pred. No. 6.3e-17;
 Matches 60; Conservative 30; Mismatches 54; Indels 23; Gaps 5;

QY 20 MTGANKSTSOVMV-----AP-NAPGYGYVITYGAFLVNDDEVYKALASK 65
 DB 26 LVASCKYDEKPPHYVKTMTPTVSTPAFTTAPDSSG-LYT--AAOLDTDACLR----- 77
 QY 66 LPSLVKEDFDSDEIKPQAAALIDBOAFLTTNQTARVLVAGHDERGSRNNSLGERRA 125
 DB 78 -QRYVYEDFDKDDVKKEQYVLGCHAKYLRNRPASHTLTIGNDERSREYNALGRRG 136
 QY 126 VAVNNYLLGKGINASVEIISFGERRIAGTNEANSQNRRELST 172
 DB 137 NSVLYSLQANGASSGQLNVVSYGEREPTCTESCSWNRNRREIIV 183

RESULT 10

0926C3 PRELIMINARY; PRT; 176 AA.

AC 0926C3; 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Probable peptidoglycan-associated lipoprotein precursor.
 GN PAL OR R02738 OR SMC02942.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;

RP STRAIN=1021;

RC MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetalle D., Puehler A., Punelle B., Ramsperger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.",
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL: AL591791; CAC47317.1; -
 DR InterPro: IPR001145; Bac_Ompa.
 DR Pfam: PF00691; Ompa.1.
 DR ProDom: PD000930; Bac_Ompa.1.
 DR Complete proteome.
 SQ SEQUENCE 176 AA; 18735 MW; A6F162CB35042268 CRC64;

Query Match 30.5%; Score 262.5; DB 16; Length 176;
 Best Local Similarity 38.5%; Pred. No. 1.1e-16;
 Matches 60; Conservative 24; Mismatches 61; Indels 11; Gaps 3;

QY 13 ALSVTFMPCANKSTSOVMVAPNAPGYGYVITYGAFLVNDDEVYKALASLPVYF 72
 DB 23 ALVWTLTALACCAKSN-----LPNDAG-LGLAGAAATPSSQDDFTV-----NVGDRIF 71
 QY 73 DFDSEIKPQAAALIDBOAFLTTNQTARVLVAGHDERGSRNNSLGERRAVAVRNTL 132
 DB 72 DFDSTSRADAQATLDRQAQWLAKYPNYGTTIEGHADERTRTYNTALARRAAATRDYL 131
 QY 133 LGKGINASVEIISFGERRIAGTNEANSQNRRA 168
 DB 132 VSRGVGNRRRTTSYKREKREAVACDDISCSQNRRA 167

RESULT 11

0809L5 PRELIMINARY; PRT; 177 AA.

AC 0809L5; 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Omp16 protein.

GN P16A OR ATU0713 OR AGR_L_2246.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

OC NCBI_TaxID=176299;

RP STRAIN=11743193;

RC MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F.Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Raymond C., Rose G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 Zhang S., Ioo H., Tao Y., Biddle P., Jung K., Krespan W., Perry M.,

OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E:
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 Stenstrom M., Winkler H.H., Kurland C.G.,
 "The genome sequence of Rickettsia prowazekii and the origin of
 mitochondria."
 RT Nature 396:133-140(1998).
 RL EMBL: AJ235273; CAI5198.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA_1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA_1.
 KW Complete proteome.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 1 25 168
 SQ SEQUENCE 155 AA; 17334 MW; D486CF6BEC40BBF CRC64;
 Query Match 29.7%; Score 256; DB 16; Length 152;
 Best Local Similarity 35.7%; Pred. No. 3.8e-16;
 Matches 56; Conservative 26; Mismatches 59; Indels 16; Gaps 2;
 OY 13 ALSVLFMTGC-ANKSTQVWVAPNAPGTGYITGVAPLVNDDEVKALASKLPSTLY 71
 Db 8 AFALPMLAGCNTTKTKTQM-----AGWHQCEETSLIKDEKHAAGNAVW 52
 OY 72 FPDFSDEIKPQAAAILDEQAQFLTNTQTAFLVAGHTDEGSRREYNMISGRRVAVKRY 131
 Db 53 FAFDSSSLSPKAKEELERQACWLSKHPEVKYVECHCEKGTREYNALGERRAAAKQF 112
 OY 132 LUGKGNASVELISFGEERPIAFGTNEAMSONRRA 168
 Db 113 LANKGIAHNRMLTISYGRKDPAMIGTTEAFSYNRR 149
 RESULT 15
 OX 044662 PRELIMINARY; PRT; 168 AA.
 AC 044662;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE OMP16 precursor (PAL protein) (Peptidoglycan-associated
 lipoprotein)
 GN PAL OR BME10340.
 OS Brucella abortus 1, and
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=35602; 29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B. abortus 1, and B. melitensis; STRAIN=544, AND 16M;
 RX MEDLINE=94341863; PubMed=8063379;
 RA Tibor A., Weynants V., Denoel P., Lichtfouse B., De Bolle X.,
 Saman E., Linet J.N., Letesson J.J.,
 "Molecular cloning, nucleotide sequence, and occurrence of a 16.5-
 kilodalton outer membrane protein of Brucella abortus with similarity
 to pal lipoproteins."
 RT Infect. Immun. 62:3633-3639(1994).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B. melitensis; STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RA Tibor A., Aidant N., Letesson J.J.,
 "The tol-pal region in Brucella encodes homologs of the Tol-Pal system
 of E. coli."
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B. melitensis; STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;

RA DelVecchio V.G., Kaparat V., Redkar R.J., Patra G., Mufer C., Los T.,
 Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 Selkov E., Elzer P.H., Haggis S., O'Callaghan D., Letesson J.-J.,
 Haselkorn R., Kyriades N., Overbeek R.,
 "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 RL EMBL: L27966; AAA59360.1;
 DR EMBL: AF358662; AAK48919.1;
 DR EMBL: AE009476; AAL51521.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA_1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA_1.
 KW Signal; Complete proteome.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 1 25 168
 SQ SEQUENCE 168 AA; 18232 MW; 17F754F05EB1D8 CRC64;
 Query Match 29.5%; Score 254; DB 16; Length 168;
 Best Local Similarity 36.3%; Pred. No. 6.5e-16;
 Matches 62; Conservative 28; Mismatches 65; Indels 16; Gaps 5;
 OY 2 MLHIQIAAALSVITFT-----GCANKSTQVWVAPNAPGTGYITGVAPLVNDDE 57
 Db 1 MRRIQ-SIARSPIALALFMSLAVAGCASKKN-----LPN-NAGDLGAGCAATPGSSQDF 53
 OY 58 TVKALASKLPSTLYPDFSDEIKPQAAAILDEQAQFLTNTQTAFLVAGHTDEGSRREYN 117
 Db 54 TV-----NVGDRIFPDFDLSLRADAQOTLSKQAWLQRYQYISITTEGADRGREYN 108
 OY 118 NSLGERRAVAVNYNYLGGKGNASVELISFGEERPIAFGTNEAMSONRRA 168
 Db 109 LALGQRAAATFEDFLASRGVPTNRMRTISYGRNBPVAVCADTCMSQNRRA 159
 RESULT 16
 OX 044157 PRELIMINARY; PRT; 152 AA.
 AC 044157;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PALA outer membrane protein.
 GN PALA
 OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Actinobacillus.
 OX NCBI_TaxID=715;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96298014; PubMed=8763621;
 RA Frey J., Kuhner P., Villiger L., Nicolet J.,
 "Cloning and characterization of an Actinobacillus pleuropneumoniae
 outer membrane protein belonging to the family of PAL lipoproteins."
 RL Res. Microbiol. 147:351-361(1996).
 DR EMBL: X89009; CA61413.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA_1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR PRODOM: PD000930; Bac_OmpA_1.
 FT CHAIN 20 152
 SQ SEQUENCE 152 AA; 16235 MW; 892C1E23E80406A7 CRC64;
 Query Match 29.4%; Score 253.5; DB 2; Length 152;
 Best Local Similarity 37.5%; Pred. No. 6.3e-16;
 Matches 63; Conservative 27; Mismatches 55; Indels 23; Gaps 5;
 OY 5 IOIAAALSVLFTMTGCANKSTQVWVAPNAPGTGYITGVAPLVNDDEVKALAS 64
 Db 8 LMTAAPA-----FVLAACSSSSDANANANANAGQFGGM-----TREDLQT 48

RA Raoult D.;

DE
reputatorem

GN HP1125.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_TaxID=210;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RA MEDLINE=9739467; PubMed=9252185;
 RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gokey J.D., Osterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 DR EMBL: AE000619; AAD08169.1; -
 DR TIGR: HP1125; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA. 1.
 DR PRINTS: PRO1021: OMPADOMAIN.
 DR Prodom: PD000930; Bac_OmpA. 1.
 KW Hypothetical protein; Lipoprotein; Complete proteome.
 SQ SEQUENCE 179 AA; 19978 MW; 5B5521E284E27BAC CRC64;

Query Match 25.0%; Score 215.5; DB 16; Length 179;
 Best Local Similarity 32.2%; Pred. No. 2.8e-12;
 Matches 56; Conservative 34; Mismatches 63; Indels 21; Gaps 5;

QY 15 SVLFPM-----TGCANKSTSGQVWVAP-NAPTYGTGYITV-----GVAPLYDND 56
 DB 5 SVFSFLVAFLLVYGCGHKMNKTAVAGDVSAKTQVATPTTEPAPEKEEPKQEPAPVVEEK 64
 QY 57 ETVALASKLPSLVYEDFSDSEIKPQAAALIDBQAFLLTNGARVAVAGHTDERSGREY 116
 DB 65 PAYD-SGTLIASYFDKYEIKESDQETLDEIVQAKENH-MQVLLGNTDDEGSSSEY 121
 QY 117 NMSIGERRAVAVNRYLLGKINGQASVEIISFGEERPIAGTNEAMSONRAEL 170
 DB 122 NOALGVKRTLSVKNALVIKVEKDMKITISFGETKPKCAOKTRECYNRRVDV 175

RESULT 21
 Q92K87 PRELIMINARY; PRT; 179 AA.
 ID Q92K87
 AC Q92K87;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Putative outer membrane protein.
 GN UHP1054.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_TaxID=85963;
 [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=9120557; PubMed=9923682;
 RA Alm R.A., Tang L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 DR EMBL: AE001533; AAD06633.1; -
 DR InterPro: IPR001145; Bac_OmpA.

DR Pfam: PF00691; OmpA. 1.
 DR PRINTS: PRO1021: OMPADOMAIN.
 DR Prodom: PD000930; Bac_OmpA. 1.
 KW Complete proteome.
 SQ SEQUENCE 179 AA; 20030 MW; F62BB619186458BD CRC64;

Query Match 25.0%; Score 215.5; DB 16; Length 179;
 Best Local Similarity 32.8%; Pred. No. 2.8e-12;
 Matches 57; Conservative 30; Mismatches 66; Indels 21; Gaps 5;

QY 15 SVLFPM-----TGCANK-----STSGQVWVAP-NAPTYGTGYITVGAAPLYDND 56
 DB 5 SVFSFLVAFLLVYGCGHKMNKTAVAGDVSAKTQVATPTTEPAPEKEEPKQEPAPVVEEK 64
 QY 57 ETVALASKLPSLVYEDFSDSEIKPQAAALIDBQAFLLTNGARVAVAGHTDERSGREY 116
 DB 65 PAYE-SGTLIASYFDKYEIKESDQETLDEIVQAKENH-MQVLLGNTDDEGSSSEY 121
 QY 117 NMSIGERRAVAVNRYLLGKINGQASVEIISFGEERPIAGTNEAMSONRAEL 170
 DB 122 NOALGVKRTLSVKNALVIKVEKDMKITISFGETKPKCAOKTRECYNRRVDV 175

RESULT 22
 Q92AW7 PRELIMINARY; PRT; 179 AA.
 ID Q92AW7
 AC Q92AW7;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Omp22.
 GN Omp22.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_TaxID=210;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KCTC0217BP.
 RA Seo W.Y., Kim J.S., Tang J.H., Yu G.J., Yum J.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U75869; AAD09577.1; -
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA. 1.
 DR PRINTS: PRO1021: OMPADOMAIN.
 DR Prodom: PD000930; Bac_OmpA. 1.
 SQ SEQUENCE 179 AA; 20024 MW; 025ADC298F50C5AC CRC64;

Query Match 24.2%; Score 208.5; DB 2; Length 179;
 Best Local Similarity 30.8%; Pred. No. 1.2e-11;
 Matches 52; Conservative 33; Mismatches 69; Indels 15; Gaps 4;

QY 14 LSVLFPMGCAKSTSGQVWVAP-NAPTYGTGYITV-----APLYDNDIFYKA 61
 DB 10 LVAFLEFAGCKAHNMDEKTVAVAGDVSAKAVASAVSTETIAQEKQEPKQEPAPVVEEK 68
 QY 62 LASKLPSLVYEDFSDSEIKPQAAALIDBQAFLLTNGARVAVAGHTDERSGREY 121
 DB 69 -SGTLIASYFDKYEIKESDQETLDEIVQAKENH-MQVLLGNTDDEGSSSEY 126
 QY 122 ERRAVAVNRYLLGKINGQASVEIISFGEERPIAGTNEAMSONRAEL 170
 DB 127 VKRTLSVKNALVIKVEKDMKITISFGETKPKCAOKTRECYNRRVDV 175

RESULT 23
 O8XWUO PRELIMINARY; PRT; 209 AA.
 ID O8XWUO
 AC O8XWUO;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Probable transmembrane protein.

OY 104 VAGHTDERSREYNMSIGERRAVARNYLKGINQASVELISPEERPIAFGNEBWS 163
 DB 95 VEGNCDEWGTDEYNQALGKRAKAVKELLAKGVNADRIKAVSGEINPCTEXTKACDA 154
 OY 164 QNRRAE 169
 DB 155 QNRRAE 160

RESULT 26

0927C5 PRELIMINARY; PRT; 192 AA.
 AC 0927C5;
 DT 01-MAY-1999 (TReMBLrel. 10, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
 DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 GN Peptidoglycan-associated lipoprotein.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Ulfertback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shida T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWT029 from USA";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL: AE001659; AND18919.1; -;
 DR EMBL: AE002265; AAF38862.1; -;
 DR EMBL: AF002547; BAA98989.1; -;
 DR TIGR: CP1091; -;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR ProDom: PD000930; Bac_OmpA; 1.
 KW Lipoprotein; Complete Proteome.
 SQ SEQUENCE 192 AA; 21925 MW; A55094A12DD50D50 CRC64;

Query Match 23.8%; Score 204.5; DB 16; Length 192;
 Best Local Similarity 38.7%; Pred. No. 3.2e-11;
 Matches 46; Conservative 24; Mismatches 46; Indels 3; Gaps 2;

OY 54 DNDDEVKALASKLPFL--VYFDEDSDEIK-PQAAIIDEQAQFLTTQTAFLVAGHTDE 110
 DB 68 DSKEKQYKSSQVAFAFRNITPATDSYTKGEEINLALINLVHMKKNKATLYIEGHTDE 127
 OY 111 RGSREYNMSIGERRAVARNYLKGINQASVELISPEERPIAFGNEBWSQNRRAE 169
 DB 128 RGAASYNIALGARRANAIEKHLRKQGISADRLSTISYKREHPLNSGNHLMWQONRRTE 186

RESULT 27

092ST9 PRELIMINARY; PRT; 221 AA.
 AC 092ST9;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 GN Hypothetical transmembrane protein SMC00354.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godtke T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Gallibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL: AL591783; CAC41711.1; -;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA; 1.
 DR ProDom: PD000930; Bac_OmpA; 1.
 DR PROSITE: PS01068; OMPA; UNKNOWN 1.
 KW Hypothetical protein; Complete Proteome.
 SQ SEQUENCE 221 AA; 22852 MW; 3FF7CBA8F69DD396 CRC64;

Query Match 23.7%; Score 204; DB 16; Length 221;
 Best Local Similarity 37.6%; Pred. No. 4.4e-11;
 Matches 47; Conservative 17; Mismatches 57; Indels 4; Gaps 1;

OY 47 TGVAPLVNDDEVKALASKLPFLSYFPEFDSIEIPQAAIIDEQAQFLTTQTAFLVAG 106
 DB 94 TGVSVTRAGDRIT---LNMPSNTTFATDRQVVPARYSTLDSVAIVLRFRNKLLIDVDG 149
 OY 107 HTDERSREYNMSIGERRAVARNYLKGINQASVELISPEERPIAFGNEBWSQNR 166
 DB 150 HTDSTGASVYNGSLERRAASVANYLASRGVDQRRMAVGYGLERPIASNAATEGVRQNR 209
 OY 167 RAELS 171
 DB 210 RVEIS 214

RESULT 28

098NA9 PRELIMINARY; PRT; 216 AA.

AC 098NA9;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 GN Hypothetical protein mlr0220.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimojo M., Sugimoto M.,
 RA Takeuchi C., Yamada W., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium

RT Mesorhizobium loti.
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP002994; BAB47852.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA.1.
 DR Hypothetical protein: complete
 KW SEQUENCE 216 AA; 22401 MW; 6FC9E5B8BAC404A CRC64;

Query Match 23.5%; Score 202; DB 16; Length 216;
 Best Local Similarity 36.3%; Pred. No. 6.5e-11;
 Matches 45; Conservative 20; Mismatches 55; Indels 4; Gaps 1;

QY 47 TGVAPLVNDDETAKALSKLPSTLYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVYAG 106
 DB 91 TGVSVTRSGDQII-----LNPMSDITFENVDDAVKGFYPLNSVALYKKRQTTVVF 146
 QY 107 HTDGRSGREYNSIGERRAVVRNYLKGKINGASVETISFGERRPIAFGTNEBWSQNR 166
 DB 147 HTDSTGCGDQHNFDLSORALAVANTLSQGYDQRRFATGTRPIASNTAAGRCQNR 206
 QY 167 RAEL 170
 DB 207 RVEI 210

RESULT 29
 OVR19 PRELIMINARY; PRT; 249 AA.

AC 08VR19;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 26.3 kDa protein.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
 OX NCBI_Taxid=34;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kirby J.R., Zusman D.R.,
 RT "A Chemotaxis System from Myxococcus xanthus Regulates Developmental
 Gene Expression Rather than Motility".
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF448145; AAL46989.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA.1.
 DR Hypothetical protein.
 KW SEQUENCE 249 AA; 26335 MW; 170AF8EFAF632855 CRC64;

Query Match 23.4%; Score 201.5; DB 2; Length 249;
 Best Local Similarity 40.2%; Pred. No. 8.8e-11;
 Matches 41; Conservative 19; Mismatches 41; Indels 1; Gaps 1;

QY 68 SLVYFPDSDDEIKPQAAAILDEQAQFLTTNQTARVYAGHTDGRSGREYNSIGERRAVA 127
 DB 146 SLPVFPFNSGLSADNQRRSLDAQCMKT-APGRVTLAHADRGTEENLQLSNRRAS 204
 QY 128 VRYNYLKGKINGASVETISFGERRPIAFGTNEBWSQNR 169
 DB 205 VKRYTLDLGVASOLGTGVGTRPVNSASSEDWASENRRVE 246

RESULT 30
 ID 046099 PRELIMINARY; PRT; 165 AA.
 AC 046099;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE OMP18 protein.
 CN PAL OR C1AD.
 OS Campylobacter jejuni.
 OS Plasmid pUMW97.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_Taxid=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M275.
 RX MEDLINE=96201603; PubMed=8613402;
 RA Konkel M.E., Mead D.J., Clepiak W. Jr.,
 RT "Cloning, sequencing, and expression of a gene from Campylobacter
 jejuni encoding a protein (Omp18) with similarity to peptidoglycan-
 associated lipoproteins".
 RT Infect. Immun. 64:1850-1853(1996).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=72DZ/92;
 RC Pawelec D.P., Jaguszyn-Krynicka K.E.,
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U47617; AAC35420.1;
 DR EMBL: AJ132802; CA10786.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR PRINTS: PR01021; OMPADOMAIN.
 DR ProDom: PD000930; Bac_OmpA.1.
 DR Plasmid: Signal.
 KW SEQUENCE 165 AA; 17890 MW; 35E67427864B9C9D CRC64;

Query Match 23.2%; Score 199.5; DB 2; Length 165;
 Best Local Similarity 33.1%; Pred. No. 7.6e-11;
 Matches 57; Conservative 26; Mismatches 70; Indels 19; Gaps 5;

QY 2 MLHQTAAALSVLTFMTGCAKSTSQVMPNAPGTGYGVY--TGVAPLVNDDET 58
 DB 4 ILSSIAAFPLVTS-----GCSTKTS-----VSGDTSVDSNRGTGSGDGMIDSK 49
 QY 59 VKALASKLPSTLYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVYAGHTDGRSGREN 117
 DB 50 ISQINDTL-GKVFYDFPKFNIRPDMQVNTNANINNEVSGVITVEGCDWGTDEYN 108
 QY 118 MSIGERRAVVRNYLKGKINGASVETISFGERRPIAFGTNEBWSQNR 169
 DB 109 QALGRKRAVKEALIQVNSDRIAVKSGETINPCTETKACDAQNRRAE 160

RESULT 31
 ID 084605 PRELIMINARY; PRT; 188 AA.
 AC 084605;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Peptidoglycan-associated lipoprotein.
 GN PAL OR C1600.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_Taxid=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/CX;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lamm C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.,
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis".
 RL Science 282:754-759(1998).
 DR EMBL: AE001330; AAC68202.1;
 DR InterPro: IPR001145; Bac_OmpA.
 DR Pfam: PF00691; OmpA.1.
 DR ProDom: PD000930; Bac_OmpA.1.

RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Homiel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
RA Molian C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cleto C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL: AE009218; AAL43703.1;
DR EMBL: AE008185; AAK8440.1;
KW Complete proteome.
SQ SEQUENCE 220 AA; 22640 MM; C3CED820F0424E2 CRC64;

Query Match 22.0%; Score 189; DB 16; Length 220;
Best Local Similarity 33.9%; Pred. No. 1.1e-09;
Matches 42; Conservative 22; Mismatches 56; Indels 4; Gaps 1;

QY 47 TGVAPLVNDETVKALASLPVLYFDPDSDEIKPQAAAILDEQAQFLTTNQTARVLYAG 106
DB 94 TGVSVTRRGD---SIVLNMPSNITFATDQDVIPFQTLDSVAIVLNKRNLLIDING 149
QY 107 HTDEGRSREYNMSLGGRRVAVRNYLLKGINQASVEITISGEERPIAFGTNEAWSONR 166
DB 150 HTDSTSLQHNQALSERRAASVANYLGARGVDORISITLGGPSPPIASNATSDGRQNR 209
QY 167 RAEL 170
DB 210 RVEV 213

RESULT 38
ID Q8XRSO PRELIMINARY; PRT; 277 AA.
AC Q8XRSO;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Probable lipoprotein.
GN RSP0761 OR RS01947.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OC NCBI_Taxid=305;
OX RN
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM1000;
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangelot S.,
RA Ariat M., Billault A., Brothier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siglier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Welzenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502(2002).
RL EMBL: AL646080; CAD17912.1;
DR InterPro: IPR001145; Bac.OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac.OmpA; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 277 AA; 28898 MM; D08C29F4498FF525 CRC64;

Query Match 22.0%; Score 189; DB 16; Length 277;
Best Local Similarity 31.3%; Pred. No. 1.5e-09;
Matches 47; Conservative 29; Mismatches 64; Indels 10; Gaps 3;

QY 25 NKSTQVMVAPNAPGTGTYITV-----GAPLVNDETVKALASLPVLYFDPDSDEI 79
DB 128 NEAQRQOLASDMATV--LGASYIKNGLRVAAVSEQLSDQTLANR---IIEFFEGSANTL 182
QY 80 KPOAAALIDEOAQFLTTNQTARVLYVGHTEGRSREYNMSLGGRRVAVRNYLLKGINQ 139
DB 183 TPGRGTLDDMAVLRNLGRKLEIYGHTEGRSRRAMNLSQARAETVKNYLLSKGAEP 242
QY 140 ASVEITISGEERPIAFGTNEAWSONRAE 169
DB 243 GTLTAIGVGPDPVAPNATDEGRSRRRIE 272

RESULT 39
ID Q8YHLL PRELIMINARY; PRT; 232 AA.
AC Q8YHLL;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Outer membrane protein.
GN BMEI0786.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_Taxid=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RC MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyprides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009519; AAL51967.1;
DR InterPro: IPR001145; Bac.OmpA.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01021; OMPADOMAIN.
DR ProDom: PD000930; Bac.OmpA; 1.
KW Complete proteome.
SQ SEQUENCE 232 AA; 24422 MM; 528DE7B4367E73FF CRC64;

Query Match 21.8%; Score 187.5; DB 16; Length 232;
Best Local Similarity 27.1%; Pred. No. 1.6e-09;
Matches 58; Conservative 25; Mismatches 84; Indels 47; Gaps 5;

QY 4 HIOIAAAAALSVLFMTGC-ANKSTQVWV-----APNAPGTGTYITG----- 48
DB 11 HFMLKGTGIALICATFLAGCTDPTTGQKVSNTGGAALGAAGVLAGLGGSSRAOR 70
QY 49 -----VAPLVNDET--VKA-----DASKLPSLYFDPDS 76
DB 71 NAVLIGAGIGALGGGLIGNVROEALRAQLOGTGVSVTRNDRITLNMPSNITFPDDQ 130
QY 77 DEIKPQAAALIDEOAQFLTTNQTARVLYVGHTEGRSREYNMSLGGRRVAVRNYLLKG 136
DB 131 DQVKSQFVPLTNSVAIVLRKFDKTLVDYIGFTDSTGSASTNOLSORRAASVASTDSOG 190
QY 137 INQASVEITISGEERPIAFGTNEAWSONRAEL 170
DB 191 IDPRFAVIVGYGASOPTASNATPEGRAQNRVREI 224

RESULT 40
Q9XCZ7

ID O9XCZ7 PRELIMINARY; PRT: 328 AA.
AC O9XCZ7;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Outer membrane protein OprF (Fragment).
GN .OPRF.
OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=316;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LWG 11197;
RX MEDLINE=99319345; PubMed=10390872;
RA Vermeiren H., Willems A., Schoofs G., de Mot R., Kelders V., Hai W.,
Vanderleyden J.;
RT "The rice inoculant strain *Alcaligenes faecalis* A15 is a nitrogen-
fixing *Pseudomonas stutzeri*."
RL Syst. Appl. Microbiol. 22:215-224(1999).
DR EMBL: AF117971; AAD39366.1; -.
DR InterPro: IPR001145; Bac_OmpA.
DR InterPro: IPR001035; MotY.
DR Pfam: PF00691; OmpA; 1.
DR PRINTS: PR01023; NAFLGMOTY.
DR PRODOM: PD000930; Bac_OmpA; 1.
DR PROSITE: PS01068; OMPA; UNKNOWN_1.
FT NON_TER 328
SQ SEQUENCE 328 AA: 35302 MW: 5FCDF79C6955090 CRC64;

Query Match 21.8%; Score 187.5; DB 2; Length 328;
Best Local Similarity 34.7%; Pred No. 2.6e-09;
Matches 42; Conservative 21; Mismatches 57; Indels 1; Gaps 1;

OY 50 APVVDNETVKALASLPVYFDPDSDEIKPQAAIIDEQAQFTTNGTARVLVAGHTD 109
||:| | : : | | | | : : : : | : : : | | | |
DB 202 APIVDTEPEPAPEVVRVELDVKFDKSRVREESYSDIKNLADFMQYPTTTVEGHTD 261
OY 110 ERGSRREYNKSLGERAVAVRNVLGK-GINQASVELISGEEPRPIAFGTNEEAWSONRRA 168
| : : | | | | | | | | : : : : | : : : | | | |
DB 262 SVGTDOYNOGLSERRAEAVRNVLVNEYGVQGNRVNSVGTGESRPVADNSTEERGQINRRV 321
OY 169 E 169
|
DB 322 E 322

Search completed: July 6, 2003, 14:06:30
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:06:35 ; Search time 35 Seconds
(without alignments)
144.593 Million cell updates/sec

Title: US-09-674-779b-2
Perfect score: 861
Sequence: 1 MMLHIQIAAAALSVLFEM.....IAFTNEEAMQNRRAELSY 172

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/laa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/laa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/laa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/laa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/laa/PCRTUS.COMB.pep:*
6: /cgn2_6/prodata/1/laa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	30.7	153	6 5173294-2	Patent No. 5173294
2	175.5	20.4	214	2 US-08-572-447C-13	Sequence 13, Appl
3	175.5	20.4	214	4 US-09-267-747-13	Sequence 13, Appl
4	174.5	20.3	338	1 US-08-210-394-1	Sequence 11, Appl
5	171.5	19.9	161	2 US-08-572-447C-11	Sequence 11, Appl
6	171.5	19.9	161	4 US-09-267-747-11	Sequence 11, Appl
7	171.5	19.9	226	4 US-08-572-447C-15	Sequence 15, Appl
8	171.5	19.9	226	4 US-09-267-747-15	Sequence 15, Appl
9	165.5	19.2	359	1 US-08-457-997B-2	Sequence 2, Appl
10	165.5	19.2	359	3 US-08-467-722A-2	Sequence 2, Appl
11	136.5	15.9	335	4 US-08-836-500A-2	Sequence 2, Appl
12	87.5	10.2	507	4 US-08-469-260A-605	Sequence 605, Appl
13	87.5	10.2	1422	4 US-08-469-260A-83	Sequence 83, Appl
14	87.5	10.2	2864	4 US-08-469-260A-394	Sequence 394, Appl
15	84	9.8	30	1 US-08-475-989-41	Sequence 41, Appl
16	84	9.8	30	3 US-08-475-985-41	Sequence 41, Appl
17	84	9.8	30	3 US-08-256-839-41	Sequence 41, Appl
18	81	9.4	964	4 US-09-556-877-177	Sequence 177, Appl
19	81	9.4	964	4 US-09-620-412C-191	Sequence 191, Appl
20	81	9.4	977	4 US-09-556-877-191	Sequence 191, Appl
21	81	9.4	977	4 US-09-620-412C-191	Sequence 191, Appl
22	79	9.2	24	1 US-08-475-985-39	Sequence 39, Appl
23	79	9.2	24	2 US-08-475-985-39	Sequence 39, Appl
24	79	9.2	24	2 US-08-256-839-39	Sequence 39, Appl
25	79	9.2	2616	6 5206163-3	Patent No. 5206163
26	76	8.8	885	3 US-09-074-579-5	Sequence 5, Appl
27	76	8.8	885	4 US-09-388-774-5	Sequence 5, Appl


```

Query Match      20.4%; Score 175.5; DB 4; Length 214;
Best Local Similarity 37.5%; Pred. No. 3.7e-13;
Matches 39; Conservative 16; Mismatches 48; Indels 1; Gaps 1;

QY 70 VYFDPDSDEIKPQAAILDEQAOPLTTNQTARVLYAGHTDGRSREYNMSLGERRAVR 129
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51 VKFDPDSKVENSYADIKNLADPMQYPSSTVTEGHDSVGTDAYNOKLSERRANAVR 110
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 NYLLGK-GINQASVEIISFGGEERPIAFGTNEEAMSONRRAEISY 172
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 111 DVLVNEGYEGGRVNAVGYESRPVADNATREGRAINRVEESH 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-08-210-394-1
; Sequence 1, Application US/08210394
; Patent No. 5770213
; GENERAL INFORMATION:
; APPLICANT: Eliotnick Dr., Gary W.
; TITLE OF INVENTION: Purified No. 5770213typable Haemophilus
; TITLE OF INVENTION: Influenzae P5 Protein as a Vaccine for No. 5770213typable
; TITLE OF INVENTION: Haemophilus Influenzae Strain
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: US
; ZIP: 07470-8426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/210,394
; FILING DATE: 07-MAR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J
; REFERENCE/DOCKET NUMBER: 32,144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201/831-3246
; TELEFAX: 201/831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-210-394-1

Query Match      20.3%; Score 174.5; DB 1; Length 338;
Best Local Similarity 28.9%; Pred. No. 1e-12;
Matches 46; Conservative 27; Mismatches 57; Indels 29; Gaps 5;

QY 33 VAPNAPTGY-----TGVIY---TGVAPLVDNDETVALSKSLPYFEPDSEIKP 81
   : || : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 184 INPNTAIHNPXIGISINAGISTRFGOGAP-----VTFESLIND--VTFAFGRANLKP 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 82 QAAAILDEQAOPLTTNQTARVLYAGHTDGRSREYNMSLGERRAVRNYLLGKGINQAS 141
   || : || : : : : : : : : : : : : : : : : : : : : : : : : :
DB 235 QAAATLDSIYGEMSGYKSAVAVAGYTDRTIGSDARNVKSQRADSVANVYFAKGVADA 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 142 VELISGGERPIAFGTNEEAMSONRRAEISY 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 295 ISATGYKANPYTGATDQVWGRMALIATLADPRRVEIA 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-08-572-447C-11

```

```

; Sequence 11, Application US/08572447C
; Patent No. 5955090
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/572,447C
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94120023.0
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: M. Paul Barker
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 05552.1395-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-572-447C-11

Query Match      19.9%; Score 171.5; DB 2; Length 161;
Best Local Similarity 37.6%; Pred. No. 7.3e-13;
Matches 38; Conservative 15; Mismatches 47; Indels 1; Gaps 1;

QY 70 VYFDPDSDEIKPQAAILDEQAOPLTTNQTARVLYAGHTDGRSREYNMSLGERRAVR 129
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 53 VKFDPDSKVENSYADIKNLADPMQYPSSTVTEGHDSVGTDAYNOKLSERRANAVR 112
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 NYLLGK-GINQASVEIISFGGEERPIAFGTNEEAMSONRRAE 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 113 DVLVNEGYEGGRVNAVGYESRPVADNATREGRAINRVEE 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-09-267-747-11
; Sequence 11, Application US/09267747
; Patent No. 6300102
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Hungerer, Klaus-Dieter
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Domdey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/572.447C
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SRD ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TOPOLGY: linear
MOLECULE TYPE: protein
US-08-572-447C-15

Query Match 19.98; Score 171.5; DB: 2; Length 226;
Best Local Similarity 37.68; Pred. No. 1,2e-12;
Matches 38; Conservative 15; Mismatches 47; Indels 1; Gaps 1;

OY 70 VFDFPDEDEKPKQAAATILDEQAQFLITNQTARVLYAGHTDGRSGSRREYNMGLGERRAVAVR 129
| | | | | : : : : : | : : : : | | | | | : : : : : | | | | |
Db 118 VKDFEDSKSKKENSXADIKNLADPFMKQYPTSTTVEGHTDVGIDAVNQLSERRAAVR 177
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
OY 130 NYLIGK-GINQASVEIISFGEERPIAFGTNEAMSONRRRE 169
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 178 DVLVNEGYEGEGRYNAVGYGSEPRVADNATAEGRAINRRAVE 218
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

RESULT 8
US-09-267-747-15
; Sequence 15, Application US/09267747
; Patent No. 6300102
; GENERAL INFORMATION:
; APPLICANT: Knapp, Bernhard
; APPLICANT: Broker, Michael
; APPLICANT: Von Specht, Bernd-Ulrich
; APPLICANT: Dondey, Horst
; TITLE OF INVENTION: Immunogenic Hybrid Protein OprF-OprI
; TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/267,747
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/572.447
; FILING DATE: 14-DEC-1995
; APPLICATION NUMBER: EP 94120023.0
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:

```



```

1 SOFTWARE: PatentIn Release #1.0,
2
3 CURRENT APPLICATION DATA:
4
5 APPLICATION NUMBER: US/08/475,988
6
7 FILING DATE: 07-JUN-1995
8
9 CLASSIFICATION: 424
10
11 PRIOR APPLICATION DATA:
12
13 APPLICATION NUMBER: US 08/256,833
14
15 FILING DATE: 03-FEB-1993

```

CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/CA93//00041
 FILING DATE: 03-FEB-1993
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9202219.3
 FILING DATE: 03-FEB-1992
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: STEWART, MICHAEL, I.
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-505 MTS:vg
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 41:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 30 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-08-475-989-41

Query Match	9.8%	Score 84;	DB 1;	Length 30;
Best Local Similarity	53.6%	Pred. No. 0.0021;		
Matches 15;	Conservative 7;	Mismatches 6;	Indels 0;	Gaps 0;

```

QY      145 ISFGEEPIAFGTNEAMSONKRAELSY 172
      :|:|:|:| | | |:|:|:|:| | |
Db      3 VSYGEEKPAVLGHDEAAYSKNRAVLAY 30

```

RESULT 16
 US-08-475-985-41
 Sequence 41, Application US/08475985
 Patent No. 5972349
 GENERAL INFORMATION:
 APPLICANT: CHONG, Pele
 APPLICANT: KANDIL, Ali
 APPLICANT: SIM, Charles
 APPLICANT: KLEIN, Michel
 TITLE OF INVENTION: Synthetic Haemophilus Influenzae
 TITLE OF INVENTION: Conjugate Vaccine
 NUMBER OF SEQUENCES: 56
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sim & McBurney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/475,985
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/256,839
 FILING DATE: 03-FEB-1993
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/CA93/00041
 FILING DATE: 03-FEB-1993
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9202219.3
 FILING DATE: 03-FEB-1992
 CLASSIFICATION: 424

```

1 ATTORNEY/AGENT INFORMATION:
2
3 NAME: STEWART, MICHAEL T.
4
5 REGISTRATION NUMBER: 24,973
6
7 REFERENCE/DOCKET NUMBER: 1038-506 MIS:vg
8
9 TELECOMMUNICATION INFORMATION:
10
11 TELEPHONE: (416) 595-1155
12
13 TELEFAX: (416) 595-1163
14
15 INFORMATION FOR SEO ID NO: 41:
16
17 SEQUENCE CHARACTERISTICS:
18
19 LENGTH: 30 amino acids
20
21 TYPE: amino acid
22
23 STRANDEDNESS: single
24
25 TOPOLOGY: linear
26
27 US-08-475-985-41

```

Query Match	9.8%	Score 84	DB 2	Length 30
Best Local Similarity	53.6%	Pred. No. 0.0021		
Matches 15; Conservative	7	Mismatches 6	Indels 0	Gaps 0

QY 145 ISFGEEPRPIAFGTNEEAWSNRRALSTY 172
:|:|:|:| | :|:|:|:|:|:|:|
Db 3 VSYGEEKPAVLGHIDEAAYSKNRRAYLAY 30

RESULT 17
US-08-256-839-41
; Sequence 41, Application US/08256839

```

; GENERAL INFORMATION:
;
; APPLICANT: CHONG, Pele
; APPLICANT: KANDIL, Ali

```

TITLE OF INVENTION: Synthetic Haemophilus Influenzae
 TITLE OF INVENTION: Conjugate Vaccine
 NUMBER OF SEQUENCES: 56
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Sim & McBurney
 STREET: Suite 701, 330 University Avenue
 CITY: Toronto
 STATE: Ontario
 COUNTRY: Canada
 ZIP: M5G 1R7
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/256,839
 FILING DATE:
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: STEWART, MICHAEL I.
 REGISTRATION NUMBER: 24,973
 REFERENCE/DOCKET NUMBER: 1038-373 MIS:jdb
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 41:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 30 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-256-839-41

Query Match:	9.88;	Score 84;	DB 3:	Length 30;
Best Local Similarity	53.68;	Pred. No. 0.0021;		
Matches	15;	Conservative	7;	Mismatches 6;
				Indels 0;
				Gaps 0;
QY	145	ISFGGRIATCTNEAMSONRRRLAY	172	
	:		:	

Db 3 VSYGEEKPAVLGHDEAAYSKNRRRAVLAY 30

RESULT 18
US-09-556-877-177

; Sequence 177, Application US/09556877

; Patent No. 6432916

; GENERAL INFORMATION:

; APPLICANT: Probst, Peter

; APPLICANT: Bhatia, Ajay

; APPLICANT: Skelky, Yasir

; APPLICANT: Fling, Steve

; APPLICANT: Maisonneuve, Jeff

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C5

; CURRENT APPLICATION NUMBER: US/09/556,877

; CURRENT FILING DATE: 2000-04-19

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; NUMBER OF SEQ ID NOS: 305

; SEQ ID NO 177

; LENGTH: 964

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-556-877-177

Query Match 9.4%; Score 81; DB 4; Length 964;

Best Local Similarity 23.6%; Pred. No. 1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFMTGCAKSTQVWVAPNAPGTGYTGAVPLVNDDET--VKALASK 65

Db 277 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 334

QY 66 LPSLYVFPDSDDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDEGSRREYNM 125

Db 335 KGAAYIDGTSN-----SKISADRAHAIFFENIYVTNANGSTSNANP-----RNA 383

QY 126 VAVRN-----YLLGKGINO-----ASVEITSGEERPIAFGTNEEA 161

Db 384 ITVASSSGEILLGASSONLIFYDPIEVSNAG-----VVSFNKKA 424

RESULT 19
US-09-620-412C-177

; Sequence 177, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 177

; LENGTH: 964

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-620-412C-177

Query Match 9.4%; Score 81; DB 4; Length 964;

Best Local Similarity 23.6%; Pred. No. 1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFMTGCAKSTQVWVAPNAPGTGYTGAVPLVNDDET--VKALASK 65

Db 277 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 334

QY 66 LPSLYVFPDSDDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDEGSRREYNM 125

Db 335 KGAAYIDGTSN-----SKISADRAHAIFFENIYVTNANGSTSNANP-----RNA 383

QY 126 VAVRN-----YLLGKGINO-----ASVEITSGEERPIAFGTNEEA 161

Db 384 ITVASSSGEILLGASSONLIFYDPIEVSNAG-----VVSFNKKA 424

RESULT 20
US-09-556-877-191

; Sequence 191, Application US/09556877

; Patent No. 6432916

; GENERAL INFORMATION:

; APPLICANT: Probst, Peter

; APPLICANT: Bhatia, Ajay

; APPLICANT: Skelky, Yasir

; APPLICANT: Fling, Steve

; APPLICANT: Maisonneuve, Jeff

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C5

; CURRENT APPLICATION NUMBER: US/09/556,877

; CURRENT FILING DATE: 2000-04-19

; NUMBER OF SEQ ID NOS: 305

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 191

; LENGTH: 977

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-556-877-191

Query Match 9.4%; Score 81; DB 4; Length 977;

Best Local Similarity 23.6%; Pred. No. 1.1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFMTGCAKSTQVWVAPNAPGTGYTGAVPLVNDDET--VKALASK 65

Db 290 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 347

QY 66 LPSLYVFPDSDDEIKPQAAAILDEQAQFLTTNQTARVLVAGHTDEGSRREYNM 125

Db 348 KGAAYIDGTSN-----SKISADRAHAIFFENIYVTNANGSTSNANP-----RNA 396

QY 126 VAVRN-----YLLGKGINO-----ASVEITSGEERPIAFGTNEEA 161

Db 397 ITVASSSGEILLGASSONLIFYDPIEVSNAG-----VVSFNKKA 437

RESULT 21
US-09-620-412C-191

; Sequence 191, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 191

; LENGTH: 977

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-620-412C-191

Query Match 9.4%; Score 81; DB 4; Length 977;

Best Local Similarity 23.6%; Pred. No. 1.1;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFMTGCAKSTQVWVAPNAPGTGYTGAVPLVNDDET--VKALASK 65

Db 290 ASDGAIKVTTRLDVTGNR--GRIFSDNITKNYGAIYAPVYTLVDNGPTFFINNIA 347

QY 66 LPSLVYFDESDLEIKPQAAAILDRCQAPLTTNQTARVLVAGHTDGRSREYNMSLGERRA 125
DB 348 KGAAYIDGTSN-----SKISADRHAIIFENENITVNTVNGSTSTANP-----RRNA 396
QY 126 VAVR-----YLLGKING-----ASVELISGGEERPIAFGNEEA 161
DB 397 ITVASSGEILLGAGSSONLIFYDPIEVSNAG-----VSVSNKEA 437

RESULT 22

US-08-475-989-39
; Sequence 39, Application US/08475989
; Patent No. 5679352
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: KANDIL, Ali
; APPLICANT: SIA, Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae
; TITLE OF INVENTION: Conjugate Vaccine
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,989
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,839
; FILING DATE: 03-FEB-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA93/00041
; FILING DATE: 03-FEB-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202219.3
; FILING DATE: 03-FEB-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-505 MIS:v9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-475-989-39

Query Match 9.2%; Score 79; DB 1; Length 24;
Best Local Similarity 62.5%; Pred. No. 0.0059;
Matches 15; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 111 RGSREYNMSLGERRAVAVRNYLLG 134
DB 1 RGTPEYNIALGRRADAVKGYLAG 24

RESULT 23

US-08-475-985-39
; Sequence 39, Application US/08475985
; Patent No. 597349
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: KANDIL, Ali
; APPLICANT: SIA, Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: Synthetic Haemophilus Influenzae
; TITLE OF INVENTION: Conjugate Vaccine
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,985
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/256,839
; FILING DATE: 03-FEB-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA93/00041
; FILING DATE: 03-FEB-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9202219.3
; FILING DATE: 03-FEB-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-506 MIS:v9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-475-985-39

Query Match 9.2%; Score 79; DB 2; Length 24;
Best Local Similarity 62.5%; Pred. No. 0.0059;
Matches 15; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 111 RGSREYNMSLGERRAVAVRNYLLG 134
DB 1 RGTPEYNIALGRRADAVKGYLAG 24

RESULT 24

US-08-236-839-39
; Sequence 39, Application US/08256839
; Patent No. 6018019
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: KANDIL, Ali
; APPLICANT: SIA, Charles
; APPLICANT: KLEIN, Michel

;; TITLE OF INVENTION: Synthetic Haemophilus Influenzae
;; TITLE OF INVENTION: Conjugate Vaccine
;; NUMBER OF SEQUENCES: 56
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sim & McBurney
;; STREET: Suite 701, 330 University Avenue
;; CITY: Toronto
;; STATE: Ontario
;; COUNTRY: Canada
;; ZIP: M5G 1R7
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/256,839
;; FILING DATE:
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEWART, MICHAEL I.
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-373 MIS:jb
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1155
;; TELEFAX: (416) 595-1163
;; INFORMATION FOR SEQ ID NO: 39:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 24 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-256-839-39
;;
Query Match 9.2%; Score 79; DB 3; Length 24;
Best Local Similarity 62.5%; Pred. No. 0.0059;
Matches 15; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 111 RGSREYMSIGERRAAVARNYLIG 134
DB 1 RGTPEYNIAGORRAADVKGVLG 24
;;
RESULT 25
5206163-3
;; Patent No. 5206163
;; APPLICANT: RENARD, ANDRE; DINI, MARTIAL, JOSEPH
;; TITLE OF INVENTION: DNA ENCODING BOVINE DIARRHEA
;; VIRUS PROTEIN
;; NUMBER OF SEQUENCES: 3
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/550,816
;; FILING DATE: 06-JUL-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 331,037
;; FILING DATE: 29-MAR-1989
;; APPLICATION NUMBER: 752,981
;; FILING DATE: 08-JUL-1985
;; SEQ ID NO: 3:
;; LENGTH: 2616
5206163-3
;;
Query Match 9.2%; Score 79; DB 6; Length 2616;
Best Local Similarity 21.8%; Pred. No. 8.5;
Matches 38; Conservative 28; Mismatches 68; Indels 40; Gaps 7;
QY 28 TSOVAVAPNAPGTGVTGAVLVNDVETVAKLASKLSLVYFEDS--DEIKPOAA 84
DB 1675 TSEVGI---TLVGRAALMTGITIVYEKTE---FNADSPESIKIGLDEGCGDPRPODH 1728
QY 85 AILDEQAQFLTTNQTARVLVAGHDDERSREYN-----MSIGERRAVA 127
DB 1729 TLADE---IHSRDERPFLVLGSRSSMSNRAKTARINICTGRPOEIRDLMAQGMVLVA 1785

QY 128 VRN-----LLGKINQASVEIISGCEERP--IAGTNEAMSQNRRAEL 170
DB 1786 LRSFNPSELVDKGFLLDRVALDALSLSPGRKQVTTATYVELLEQEQVEI 1839
;;
RESULT 26
US-09-074-579-5
;; Sequence 5, Application US/09074579
;; Patent No. 6001596
;; GENERAL INFORMATION:
;; APPLICANT: Hillman, Jennifer L.
;; APPLICANT: Guegler, Karl J.
;; APPLICANT: Patterson, Chandra
;; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
;; NUMBER OF SEQUENCES: 5
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Incyte Pharmaceuticals, Inc.
;; STREET: 3174 Porter Drive
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94304
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/074,579
;; FILING DATE: HERewith
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cerione, Michael C
;; REGISTRATION NUMBER: 39,132
;; REFERENCE/DOCKET NUMBER: PF-0505 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-855-0555
;; TELEFAX: 650-845-4166
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 885 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: GENEBANK
;; CLONE: g1288563
US-09-074-579-5
;;
Query Match 8.8%; Score 76; DB 3; Length 885;
Best Local Similarity 25.5%; Pred. No. 3.7;
Matches 14; Conservative 16; Mismatches 25; Indels 0; Gaps 0;
QY 71 YFDESDETKPOAAAILDEQAQFLTTNQTARVLVAGHDDERSREYMSIGERRA 125
DB 177 HFEIEVDIFEPQGISMDAEASFITNDLIGSALTSFSGKGHVSKPSLDQORS 231
;;
RESULT 27
US-09-388-774-5
;; Sequence 5, Application US/09388774
;; Patent No. 6228991
;; GENERAL INFORMATION:
;; APPLICANT: Hillman, Jennifer L.
;; APPLICANT: Guegler, Karl J.
;; APPLICANT: Patterson, Chandra
;; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE

RESULT 28
 US-08-762-106-9
 ; Sequence 9, Application US/08762106
 ; Patent No. 5948677
 ; GENERAL INFORMATION:
 ; APPLICANT: Jarvik, Jonathan W.
 ; TITLE OF INVENTION: READING FRAME INDEPENDENT EPIITOP
 ; TITLE OF INVENTION: TAGGING
 ; NUMBER OF SEQUENCES: 47
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harris Brotman
 ; STREET: 202 Coast Blvd., Suite 111
 ; CITY: La Jolla
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/762,106
 ; FILING DATE: 09-DEC-1996

```

1      RESULT 29
2      US-09-320-774-9
3      ; Sequence 9, Application US/09320774
4      ; Patent No. 6265345
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Jarvik, Jonathan W.
7      ; TITLE OF INVENTION: READING FRAME INDEPENDENT EPIPTOPE
8      ; TITLE OF INVENTION: TAGGING
9      ; NUMBER OF SEQUENCES: 47
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Harris Brotman
12     ; STREET: 202 Coast Blvd., Suite 111
13     ; CITY: La Jolla
14     ; STATE: California
15     ; COUNTRY: US
16     ; ZIP: 92037
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: Patentln Release #1.0, Version #1.25
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/09/320,774
24     ; FILING DATE:
25     ; CLASSIFICATION:
26     ; PRIOR APPLICATION DATA:
27     ; APPLICATION NUMBER: US 08/762,106
28     ; FILING DATE:
29     ; ATTORNEY/AGENT INFORMATION:
30     ; NAME: Brotman, Harris F.
31     ; REGISTRATION NUMBER: 35,461
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: (619) 654-2428
34     ; INFORMATION FOR SEQ ID NO: 9:
35     ; SEQUENCE CHARACTERISTICS:
36     ; LENGTH: 527 amino acids
37     ; TYPE: amino acid
38     ; TOPOLOGY: linear
39     ; MOLECULE TYPE: protein
40     ; FRAGMENT TYPE: internal
41     ; US-09-320-774-9

```

Query Match 8.5%; Score 73.5; DB 4; Length 527;
Best Local Similarity 21.2%; Pred. No. 3.3;
Matches 33; Conservative 21; Mismatches 77; Indels 25; Gaps 5;

QY 38 PTGTGVTYTVAVLVNDEYVRLASKPLSLV-----YFDPDSP-----EIK 80
DB 59 PADATQOTKGPSPASTKRTTRLPPTPEAVDGRHIDIQTVYLELTIVPEADTS 118
QY 81 POAAALIDE-OAQFLTNOARVLVAGHTDERGSREYNMSLGER-----RAVAVRNLLG 134
DB 119 TQTDALFDRPPPLPPLVPOKKTGTDALITQTTDRQTRQSGIENGDLDFCFEVEPILEVAVG 178
QY 135 KGINQASVEIISFGEEPIAFGTNEAMSQNRRAEL 170
DB 179 KVLQGLMEVLE--EELIAMAMRAHQEHFEIIRNAEL 212

RESULT 30
US-09-345-882-29
; Sequence 29, Application US/09345882
; Patent No. 639373
; GENERAL INFORMATION:
; APPLICANT: Bouquelere, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; TITLE OF INVENTION: A POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 29
; LENGTH: 1312
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 294..296
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 432..434
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 755..757
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 856..858
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 859..861
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 910..912
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1151..1153
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: 1226..1228
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION

LOCATION: 102..105
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 663..666
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 808..811
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 885..888
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 17..19
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 31..33
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 41..43
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 100..102
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 140..142
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 216..218
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 471..473
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 507..509
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 531..533
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 591..593
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 656..658
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 801..803
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 812..814
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 815..817
; OTHER INFORMATION: potential
; FEATURE:
; NAME/KEY: PHOSPHORYLATION
; LOCATION: 876..878

OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 888..890
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 939..941
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1060..1062
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1128..1130
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1129..1131
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1135..1137
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1181..1183
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1208..1210
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 1249..1251
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 47..50
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 126..129
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 157..160
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 158..161
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 159..162
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 216..219
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 274..277
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 276..279
OTHER INFORMATION: potential
FEATURE:
NAME/KEY: PHOSPHORYLATION
LOCATION: 295..298
OTHER INFORMATION: potential

FEATURE:
Query Match 8.5%; Score 73; DB 4; Length 1312;
Best Local Similarity 22.1%; Pred. NO. 16;
Matches 34; Conservative 22; Mismatches 60; Indels 38; Gaps 5;
QY 27 STSQVMAPNAPTGTGYITGV-----PLVNDDETVALASKLPSLYPFDSD 76
DB 1016 SNSVLTPTPTPESPSTVTGSRQSSVTSEPLAPNDEVRISKSTDTIEVDVA 1075
QY 77 DEIKP-----QAAIIDEQAQFLTNO-----TARVLY-----AGHDERGSR 115
DB 1076 GELDIOSEGNSSPAGPDAVSSSSSNOPEHPEKACTGKRVKAQGGSSSKRKORS 1135
QY 116 YMSIGERRAVAVRNLYLKGINGQASVEIISFGE 149
DB 1136 H-----KATVNNKKKKGTNSDSEELSGE 1162

RESULT 31
US-08-790-912-3
Sequence 3, Application US/08790912
Patent No. 5976542
GENERAL INFORMATION:
APPLICANT: Weiser, Jeffrey N.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
OF STREPTOCOCCUS PNEUMONIAE INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,912
FILING DATE: 29-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 8.5%; Score 73; DB 2; Length 1964;
Best Local Similarity 23.6%; Pred. NO. 29;
Matches 42; Conservative 19; Mismatches 57; Indels 60; Gaps 8;
QY 17 LTFMTGANKSTSQVMAPNAPTGTGYITGVAPLVNDDETVALASKLPSLYPFDSD 76
DB 1195 LKYOTDFAN-----LPEYNLGNLTGLTYPNGLVDRDSIVKEV--LPELQKLDYOS 1243

```

Sequence 2. Application US/08960756
Patent No. 5866422
GENERAL INFORMATION:
APPLICANT: WAYNE, JAY
APPLICANT: XU, SHUANG-YONG
TITLE OF INVENTION: METHOD FOR CLONING AND
TITLE OF INVENTION: PRODUCING THE Tsp45I RESTRICTION ENDONUCLEASE IN E. COLI
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESSES:
ADDRESS: New England Biolabs, Inc.
STREET: 32 Tozer Road
CITY: Beverly
STATE: MA
COUNTRY: US
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,756
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Gregory D
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-128
TELECOMMUNICATION INFORMATION:
TELEPHONE: 978-927-5054
TELEFAX: 978-927-1705
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 413 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-960-756-2

Query Match      8 4%; Score 72.5; DB 2; Length 413;
Best Local Similarity 23.6%; Pred No. 3;
Matches 34; Conservative 19; Mismatches 72; Indels 19; Gaps 5;

Oy   38 PTGTTG---VY-----TGAAPLVNDDETAKALASLPILVFDFDSDETKPPAAAILD 88
      ||| | : : : || | | | : | | | |
Db   87 PGCAAGVGLVLYIDPYGTGGQFLVGGDETDHVAIVSQPKGOLATDPTLDGPQVEFLR 146
      || | : : : | : | | | | : | : | |

Oy   89 EQAOFLLTNQTARVLVAGHTDDSGSREYMSL---GERRAVAVNRYLLGKIGIQASVEI 144
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   147 ERLLLEIMADSLFVHIDEYGEVALLIDVEFGRRNY---NHIAIASPNKNMR 203
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy   145 ISFEGERPIAF--GTFNEAMSON 165
      :|| :: : | : : :
Db   204 KAFSGOKDMILVYSKTRDYWMNS 227

```

ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/920,281C
FILING DATE: 13-AUG-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-920-281C-2

Query Match 8.4%; Score 72.5; DB 1; Length 2431;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 26; Conservative 20; Mismatches 36; Indels 31; Gaps 5;

QY 24 ANKSTSQVAVNAPADTGYTGVYTGAPLVNDENVKALASKLPSTLYFPDSDSEIKRQA 83
Db 1756 ASRAERPVAPRKRPT-----PAPRT-----AFRKKLP-LTFGDFDEHVDALA 1798
QY 84 AAI-----LDEQAQFLTTNOTARVAVAGHTDGRSGREYMSIGERRAY 126
Db 1799 SGITFGDDVDVLRIGRAGAYTFSSDTG---SGHLQCKSVQHNLQCAQLDAV 1847

RESULT 35
US-08-466-277-2
Sequence 2, Application US/08466277
Patent No. 6190666
GENERAL INFORMATION:
APPLICANT: Garoff, Henrik
TITLE OF INVENTION: DNA Expression Systems Based on
Alphaviruses
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,277
FILING DATE: 06-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/920,281
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-466-277-2

Query Match 8.4%; Score 72.5; DB 4; Length 2431;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 26; Conservative 20; Mismatches 36; Indels 31; Gaps 5;

QY 24 ANKSTSQVAVNAPADTGYTGVYTGAPLVNDENVKALASKLPSTLYFPDSDSEIKRQA 83
Db 1756 ASRAERPVAPRKRPT-----PAPRT-----AFRKKLP-LTFGDFDEHVDALA 1798
QY 84 AAI-----LDEQAQFLTTNOTARVAVAGHTDGRSGREYMSIGERRAY 126
Db 1799 SGITFGDDVDVLRIGRAGAYTFSSDTG---SGHLQCKSVQHNLQCAQLDAV 1847

RESULT 36
US-08-110-786A-8
Sequence 8, Application US/08110786A
Patent No. 5443966
GENERAL INFORMATION:
APPLICANT: FAIRWEATHER, Neil Fraser
TITLE OF INVENTION: Expression of tetanus toxin fragment C
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye P.C.
STREET: 1100 No. 5443966th Gleebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/110,786A
FILING DATE: 23-AUG-1993 1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/777,337
FILING DATE: 29-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB90/00943
FILING DATE: 20-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8914122.0
FILING DATE: 20 June 1989
ATTORNEY/AGENT INFORMATION:
NAME: Mary J. Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 117-134
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 8:


```

1 COUNTRY: United States of America
2
3 ZIP: 60606
4
5 COMPUTER READABLE FORM:
6
7 MEDIUM TYPE: Floppy disk
8
9 COMPUTER: IBM PC compatible
10
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12
13 SOFTWARE: Patent In Release #1.0, Version #1.25
14
15 CURRENT APPLICATION DATA:
16
17 APPLICATION NUMBER: US/08/854,585
18
19 FILING DATE:
20
21 PRIOR APPLICATION DATA:
22
23 APPLICATION NUMBER: US/08/237,940
24
25 FILING DATE:
26
27 ATTORNEY/AGENT INFORMATION:
28
29 NAME: Borun, Michael F.
30
31 REGISTRATION NUMBER: 25,447
32
33 REFERENCE/DOCKET NUMBER: 27666/31954
34
35 TELECOMMUNICATION INFORMATION:
36
37 TELEPHONE: 312-474-6300
38
39 TELEFAX: 312-474-0448
40
41 INFORMATION FOR SEQ ID NO: 2:
42
43 SEQUENCE CHARACTERISTICS:
44
45 LENGTH: 1337 amino acids
46
47 TYPE: amino acid
48
49 TOPOLOGY: linear
50
51 MOLECULE TYPE: protein
52
53 US-08-854-585-2

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-0595-05512-2

Query Match          8.0%; Score 69; DB 5; Length 1337;
Best Local Similarity 21.8%; Pred. No. 49;
Matches 24; Conservative 26; Mismatches 46; Indels 14; Gaps 4;

QY      21  TGCANKSTSQVWVAPNAPGTGYGVLYTGVAPLVNDETVKALASKIPSLVYEPDPSDEIK 80
      : : : : : | : : : : : | : : : : : | : : : : : |
Db      321  SGGQSDTEVLVLVGLPEGRYRNATVYSGAANGTCGP--CAIEFRINALQVDEIVYANVIS 378

QY      81  PQAAAIL----DEQAQFLITNQTARVLYVAGHTDERGSGREYNMSLGERRAV 126
      : : : : : | : : : : : | : : : : : | : : : : : |
Db      379  ATSLTLIMKVSDNES---SSNYTKIHVAGETDSS-----NLNVSEPRAY 420

Search completed: July 6, 2003, 14:08:12
Job time : 39 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 6, 2003, 14:10:35 ; Search time 307 Seconds
(without alignments)
64.426 Million cell updates/sec

Title: US-09-674-779b-2
Perfect score: 861
Sequence: 1 MMHIOIAAALSVLEFM.....IAPGTNEAMSONRAELSY 172

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 440863 seqs, 114992915 residues

Total number of hits satisfying chosen parameters: 440863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries:

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	204.5	23.8	192	10	US-09-747-948-2
2	176.5	20.5	353	9	US-10-203-942-9
3	149	17.3	226	9	US-10-156-761-10782
4	136.5	15.9	344	9	US-10-159-953-2
5	107	12.4	380	9	US-09-998-279-26
6	107	12.4	385	9	US-09-998-279-24
7	93.5	10.9	237	9	US-10-010-160-12
8	87.5	10.2	507	8	US-08-424-5508-605
9	87.5	10.2	1422	8	US-08-424-5508-83
10	87.5	10.2	2864	8	US-08-424-5508-394
11	87.5	10.2	2864	10	US-09-742-659-2
12	87.5	10.2	2865	10	US-09-742-659-6
13	82	9.5	871	10	US-09-886-468-21
14	81	9.4	964	10	US-09-841-132-177
15	81	9.4	977	10	US-09-841-132-191
16	80	9.3	361	12	US-10-007-693-74
17	80	9.3	420	9	US-10-156-761-9543
18	77.5	9.0	877	9	US-10-156-761-13758
19	76.5	8.9	438	10	US-09-815-242-14076

20	76	8.8	885	10	US-09-828-423-5	Sequence 5, Appl1
21	75.5	8.8	155	9	US-10-156-761-13516	Sequence 1316, A
22	75.5	8.8	794	10	US-09-815-242-5697	Sequence 5697, Ap
23	75.5	8.8	802	10	US-09-815-242-12668	Sequence 12668, A
24	74.5	8.7	1330	9	US-10-156-761-10510	Sequence 10510, A
25	74	8.6	1177	9	US-10-128-714-3493	Sequence 3493, Ap
26	74	8.6	1179	9	US-10-128-714-8493	Sequence 8493, Ap
27	73.5	8.5	569	9	US-09-738-626-6179	Sequence 6179, Ap
28	73.5	8.5	928	10	US-09-815-242-10417	Sequence 10417, A
29	73	8.5	1312	9	US-10-071-179-20	Sequence 29, Appl1
30	73	8.5	1491	10	US-09-815-242-5568	Sequence 5568, Ap
31	73	8.5	1502	10	US-09-815-242-12162	Sequence 12162, A
32	73	8.5	1520	9	US-09-738-626-4444	Sequence 4444, Ap
33	72.5	8.4	851	9	US-10-156-761-99333	Sequence 9933, Ap
34	72.5	8.4	2431	10	US-09-901-106-2	Sequence 2, Appl1
35	71.5	8.3	441	9	US-10-156-761-13680	Sequence 13680, A
36	71.5	8.3	507	9	US-10-156-761-14476	Sequence 14476, A
37	71.5	8.3	2862	10	US-09-742-659-5	Sequence 5, Appl1
38	71	8.2	248	10	US-09-925-301-908	Sequence 908, Ap
39	71	8.2	259	9	US-09-996-634-133	Sequence 133, Ap
40	71	8.2	259	9	US-09-997-181-133	Sequence 133, Ap
41	71	8.2	259	9	US-09-997-182-133	Sequence 133, Ap
42	71	8.2	1352	9	US-10-156-761-9867	Sequence 9867, Ap
43	70.5	8.2	199	9	US-09-738-626-5191	Sequence 5191, Ap
44	70.5	8.2	275	9	US-10-156-761-12256	Sequence 12256, A
45	70.5	8.2	299	9	US-10-156-761-12949	Sequence 12949, A
46	70.5	8.2	550	9	US-09-906-419-41	Sequence 41, Appl1
47	70.5	8.2	550	9	US-10-119-136-41	Sequence 41, Appl1
48	70.5	8.2	568	9	US-10-136-960-8	Sequence 8, Appl1
49	70	8.1	336	10	US-09-815-242-11899	Sequence 11899, A
50	70	8.1	501	9	US-09-738-626-6084	Sequence 6084, Ap
51	69.5	8.1	402	9	US-09-712-363-147	Sequence 147, Ap
52	69.5	8.1	645	8	US-08-834-666A-18	Sequence 18, Appl1
53	69	8.0	350	9	US-10-169-046-24	Sequence 24, Appl1
54	69	8.0	1215	10	US-09-815-242-5908	Sequence 5908, Ap
55	69	8.0	1267	10	US-09-815-242-13113	Sequence 13113, A
56	69	8.0	2337	10	US-09-815-242-5834	Sequence 5834, Ap
57	69	8.0	6281	10	US-09-815-242-12996	Sequence 12996, A
58	68.5	8.0	463	9	US-09-816-467-2	Sequence 2, Appl1
59	68.5	8.0	867	9	US-10-177-744A-11	Sequence 11, Appl1
60	68.5	8.0	903	9	US-09-746-783-142	Sequence 142, Ap
61	68	7.9	389	9	US-10-156-761-13369	Sequence 13369, A
62	68	7.9	438	10	US-10-156-761-9344	Sequence 9344, Ap
63	68	7.9	568	10	US-09-815-242-5140	Sequence 5140, Ap
64	67.5	7.8	327	9	US-10-156-761-11941	Sequence 11941, A
65	67.5	7.8	327	9	US-10-156-761-13045	Sequence 13045, A
66	67.5	7.8	343	9	US-10-156-761-13124	Sequence 13124, A
67	67.5	7.8	436	10	US-09-815-242-13218	Sequence 13218, A
68	67.5	7.8	438	10	US-09-741-669-452	Sequence 452, Ap
69	67.5	7.8	438	10	US-09-815-242-10029	Sequence 10029, A
70	67	7.8	223	9	US-09-892-398-22	Sequence 22, Appl1
71	67	7.8	471	9	US-10-156-761-10494	Sequence 10494, A
72	67	7.8	750	9	US-10-156-761-12865	Sequence 12865, A
73	67	7.8	760	9	US-09-892-398-29	Sequence 29, Appl1
74	67	7.8	761	9	US-09-892-398-1	Sequence 1, Appl1
75	67	7.8	1325	9	US-10-154-452-6	Sequence 6, Appl1
76	67	7.8	1437	9	US-10-154-452-2	Sequence 2, Appl1
77	67	7.8	1437	9	US-09-934-421A-6	Sequence 6, Appl1
78	67	7.8	1437	9	US-10-162-012-38	Sequence 38, Appl1
79	66.5	7.7	132	10	US-09-925-297-779	Sequence 779, Ap
80	66.5	7.7	193	9	US-09-738-626-3874	Sequence 3874, Ap
81	66.5	7.7	314	9	US-10-156-761-13748	Sequence 13748, A
82	66.5	7.7	797	9	US-10-156-761-10907	Sequence 10907, A
83	66.5	7.7	1187	9	US-10-223-070-23	Sequence 23, Appl1
84	66	7.7	421	9	US-09-738-626-6362	Sequence 6362, Ap
85	66	7.7	423	9	US-10-156-761-12135	Sequence 12135, A
86	66	7.7	565	10	US-10-124-800-26	Sequence 26, Appl1
87	66	7.7	510	10	US-09-765-272-218	Sequence 218, Ap
88	66	7.7	998	9	US-10-101-464A-931	Sequence 931, Ap
89	66	7.7	1622	9	US-10-331-061-72	Sequence 72, Appl1
90	66	7.7	2059	9	US-10-124-800-4	Sequence 4, Appl1
91	66	7.7	5877	9	US-10-142-515-11	Sequence 11, Appl1
92	66	7.7	5935	9	US-10-243-243A-8	Sequence 8, Appl1

Query Match	10.9%	Score 93.5	DB 9	Length 237
Best Local Similarity	23.48%	Pred. No. 0.058		
Matches 36	Conservative 29	Mismatches 70	Indels 19	Gaps 6
24	ANKSQVAVAPNAFTGTYGTYLYTGAVADLVNDETAKALASKLPISLYFEDESDETKPOA	83		
86	ANK- LKKMLMAAALIPQATGISADVGVL- - -RVNSNTFFFGATATLPECKRYKMGTV	140		

OY 84 AAILDEQAOLTTNOTARVLVAGHTD-----ERGSR-EYNMSLGERRAVAVNYLLGKGIN 138
Db 141 LAVLRVYLYL-----VIRGHADIGCEITKGSFSPASNNELSGARAAQAQVLYVEHGK 192

OY 139 QASVEIISFGERRIACGTNEAMSQNRRELST 172
Db 193 ASRIRSVGYADTRPLE-PSSEGSTKNRRIEFYF 225

RESULT 8

US-08-424-550B-605
; Sequence 605, Application US/08424550B
; Patent No. US20020119447A1

GENERAL INFORMATION:

APPLICANT: JOHN N. SIMONS
APPLICANT: TAMT J. PILOT-MATTIAS
APPLICANT: GEORGE J. DAMSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUERHOFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUTIK
APPLICANT: ISA K. MUSHAMMAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550B

FILING DATE:
CLASSIFICATION: 435435
ATTORNEY/AGENT INFORMATION:
NAME: POREMBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 605:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-424-550B-605

Query Match 10.2%; Score 87.5; DB 8; Length 507;
Best Local Similarity 28.4%; Pred. No. 0.75;
Matches 38; Conservative 17; Mismatches 48; Indels 31; Gaps 6;

OY 3 LHIOIAAAAAALSVTFPMTCANKSTSQVAVNAPPGYVITGVAVPLVNDDEYKAL 62
Db 360 LAVGVVAAAYLAIDFGATCVRRCWSTISV-----VAPVDEEIVEEC 409

OY 63 ASKLPISLVYFDDSEIDEPQAAAILDEQAOLTTNOTARVLVAGHTDERGSRREYNMSLGE 122
Db 410 ASFLP-----LEAVVAAL-DKLSKSTITT-----TSPFTLETALEKLNFTLGP 450

OY 123 RRA--VAVRNYLLG 134

Db 451 HAATLALIEYCCG 464

RESULT 9

US-08-424-550B-83
; Sequence 83, Application US/08424550B
; Patent No. US20020119447A1

GENERAL INFORMATION:

APPLICANT: JOHN N. SIMONS
APPLICANT: TAMT J. PILOT-MATTIAS
APPLICANT: GEORGE J. DAMSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUERHOFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUTIK
APPLICANT: ISA K. MUSHAMMAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550B

FILING DATE:
CLASSIFICATION: 435435
ATTORNEY/AGENT INFORMATION:
NAME: POREMBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 1422 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-424-550B-83

Query Match 10.2%; Score 87.5; DB 8; Length 1422;
Best Local Similarity 28.4%; Pred. No. 3.2;
Matches 38; Conservative 17; Mismatches 48; Indels 31; Gaps 6;

OY 3 LHIOIAAAAAALSVTFPMTCANKSTSQVAVNAPPGYVITGVAVPLVNDDEYKAL 62
Db 562 LAVGVVAAAYLAIDFGATCVRRCWSTISV-----VAPVDEEIVEEC 611

OY 63 ASKLPISLVYFDDSEIDEPQAAAILDEQAOLTTNOTARVLVAGHTDERGSRREYNMSLGE 122
Db 612 ASFLP-----LEAVVAAL-DKLSKSTITT-----TSPFTLETALEKLNFTLGP 652

OY 123 RRA--VAVRNYLLG 134
Db 653 HAATLALIEYCCG 666

RESULT 10
US-08-424-550B-394

Db 1616 ASFP-----LEAMVAI-DKISTIT-----ISPFLEALAKLNFPUGP 1656
QY 123 RRA--VAVRNYLLG 134
Db 1657 HAATILALIEYCCG 1670

RESULT 13

US-09-886-468-21
; Sequence 21, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pasteur Limited
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there
; FILE REFERENCE: 77813-5
; CURRENT APPLICATION NUMBER: US/09/886,468
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/113,280
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,281
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,282
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,283
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,284
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,285
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113,385
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114,050
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,056
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,057
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,058
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,059
; PRIOR FILING DATE: 1998-12-28
; PRIOR APPLICATION NUMBER: 60/114,061
; PRIOR FILING DATE: 1998-12-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-886-468-21

Query Match 9.5%; Score 82; DB 10; Length 871;

Best Local Similarity 24.0%; Pred. No. 6.2; Indels 34; Gaps 7;

Matches 42; Conservative 18; Mismatches 81;

QY 15 SVLTFM-----TGCAKSTSOVWVAPNAPGTGYTGVAPLVNDDEYKALASKL 66
Db 426 AILAFIDSGSVSDKTGTSTANNQEVSLTNAATVSGAIVATKCTLTGNS-----476
QY 67 PSLVYFDPDSDEIKPQAAAILDEQAOFLTTNQTARVLYAGHTDERGSRREY--NMSL-GE 122
Db 477 -----LTFDGNNTAGTSGAIVTETEDFTLTGTVTFSTNTAKTGALYSKGNNSLSGN 531
QY 123 RRAVAVRNYLLGKGINASVE-----IISGEEPIAFGTNEAM--SQNRRAELS 171
Db 532 TNLFSGNKATGPNSSANQEGCGGAILSLFLESASVS--TKGLWIEDNENVSLS 584

RESULT 14

US-09-841-132-177
; Sequence 177, Application US/09841132
; Patent No. US20020061848A1

; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO: 177
; LENGTH: 964
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-177

Query Match 9.4%; Score 81; DB 10; Length 964;

Best Local Similarity 23.6%; Pred. No. 9.2;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFPMTCANKSTSOVWVAPNAPGTGYTGVAPLVNDDEY--VKALASK 65
Db 277 ASDGAIVKVTTRLDVYGNR--GRIFSDNITKNTGALIVAVTVLVNCGPTTFINNITANN 334
QY 66 LPSLVYFDPDSDEIKPQAAAILDEQAOFLTTNQTARVLYAGHTDERGSRREYNSLIGERRA 125
Db 335 KGGATIIDGTN-----SKISADRHAIIFENENIYNTVNTNAGTSTANP-----RNA 383
QY 126 VAVRN-----YLLGKGINO-----ASVELISGEEPIAFGTNEA 161
Db 384 ITVASSSGEILLGAGSSQNLIFYPDIEVSNAG---VVSFNKEA 424

RESULT 15

US-09-841-132-191
; Sequence 191, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 191
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-191

Query Match 9.4%; Score 81; DB 10; Length 977;

Best Local Similarity 23.6%; Pred. No. 9.4;

Matches 39; Conservative 23; Mismatches 75; Indels 28; Gaps 7;

QY 8 AAAAAALSVLFPMTCANKSTSOVWVAPNAPGTGYTGVAPLVNDDEY--VKALASK 65
Db 290 ASDGAIVKVTTRLDVYGNR--GRIFSDNITKNTGALIVAVTVLVNCGPTTFINNITANN 347
QY 66 LPSLVYFDPDSDEIKPQAAAILDEQAOFLTTNQTARVLYAGHTDERGSRREYNSLIGERRA 125
Db 348 KGGATIIDGTN-----SKISADRHAIIFENENIYNTVNTNAGTSTANP-----RNA 396
QY 126 VAVRN-----YLLGKGINO-----ASVELISGEEPIAFGTNEA 161
Db 397 ITVASSSGEILLGAGSSQNLIFYPDIEVSNAG---VVSFNKEA 437

RESULT 16

US-10-007-693-74
; Sequence 74, Application US/10007693
; Patent No. US2002014676A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.515C2
; CURRENT APPLICATION NUMBER: US/10/007,693
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 157
; SEQ ID NO 74
; LENGTH: 361
; TYPE: PRF
; ORGANISM: Chlamydia pneumoniae
US-10-007-693-74

Query Match 9.3%; Score 80; DB 12; Length 361;
Best Local Similarity 24.5%; Pred. No. 2.9;
Matches 37; Conservative 27; Mismatches 43; Indels 44; Gaps 8;

QY 12 AALSVLTGTCANKSTQVWAPNA-PTG-----YTGVTYGVAPLVNDETVALAKSL 66
DB 158 SALHNVLPPLGEMNTEVRAIAQAALPTAEKRDSTGICGKRFKFELE--KFLPKKT 215
QY 67 PSLVYFDEPDSDEIKPQAAAIIDEQAFLLTNOTARVLVAAGHDESGREYMSIGERRAV 126
DB 216 GAVI--DMDTKETVQ-----HOGAHY--TIGORHGL 244
QY 127 AVR-----YLLGKINGASVEIISFGEERP 152
DB 245 DLGSEKPCYVVGKNIENSIYIVR-GE DHP 274

RESULT 17
US-10-156-761-9543
; Sequence 9543, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9543
; LENGTH: 420
; TYPE: PRF
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9543

Query Match 9.3%; Score 80; DB 9; Length 420;
Best Local Similarity 23.4%; Pred. No. 3.6;
Matches 43; Conservative 33; Mismatches 84; Indels 24; Gaps 6;

QY 6 QIAAAALSVLTGTCANKSTQVWAPNAPGTGYITGVAPLVNDETVALAKSL 62
DB 172 EIAAGVGLTIDTLGVNSKLSKOLSCIAEA---TGTGTSVHKEDLDKYNQGLVDR 227
QY 63 -ASKLPVLVYFDEPDSDEIKPQA--AAIIDEQAQF-----LTTNOTARVLVAAGHDE 110
DB 228 AADKVVIVAVDGAADCAKAPLNLGSLFTDRAEFAQHRWYVAVKPGGELRASVLSADR 287

QY 111 RGSREYNMSIGERRAVARN--YLLGKINGASVEIISGERRPIAFGTNEAMSONRRA 168
DB 288 QVNPDIQVLL---RAVTHNRREIYRGEAGTGTVDVSTGLRYPKAESODEDATFTVCL 344
QY 169 ELVY 172
DB 345 QVSH 348

RESULT 18
US-10-156-761-13758
; Sequence 13758, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13758
; LENGTH: 877
; TYPE: PRF
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13758

Query Match 9.0%; Score 77.5; DB 9; Length 877;
Best Local Similarity 24.1%; Pred. No. 19;
Matches 53; Conservative 23; Mismatches 87; Indels 57; Gaps 8;

QY 10 AALSVLTGTCANKSTQVWAPNAPT--GYTGYITGVAPLVNDETVALAKSL 64
DB 268 APAAYODAVLSPGAGRRVVLATSAVSSSLVPGVRVYVDCGLAREPRVDHARLSLTYL 327
QY 65 -----KLPSLVF-----DPDSDEIKP-----QAAAIIDE 89
DB 328 RASQAGRORAGRAGREAPGTYYRCWAEBADARLPFPSPFEIKVADLTAFAALQACWGP 387
QY 90 QAGFLT-----TNOTARVLV-----AGHTDESGREYMSIGERRAVAVARNYLL 133
DB 388 EASGLALLDPPGGAAMAAASVLAAGAVDSAGRATERGVMSRLGLHPRLARALLDAAP 447
QY 134 GKGINGASVEIISFGEERPPIAFGTN--EAMSONRRAELVY 172
DB 448 EVGADRAAEVVALISEEPREYGGDLAAAMRAARRGDAV 487

RESULT 19
US-09-815-242-14076
; Sequence 14076, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes

FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 14076
LENGTH: 438
TYPE: PRT
ORGANISM: Salmonella typhi
US-09-815-242-14076

Query Match 8.9%; Score 76.5; DB 10; Length 438;
Best Local Similarity 22.2%; Pred. No. 9;
Matches 41; Conservative 29; Mismatches 78; Indels 37; Gaps 6;

QY 2 MLHIIQIAAAA-----ALSVLTMTGCAKSTQVAVNAPNPTGYGVYI 46
DB 77 LAHPSLSAAASAGVEIYDIEFCREAOPIVAITGSKSTVTIIVGEMAKAGVNV-- 134
OY 47 TGVAPLVNDDETVALKSLPSLYVPEDESDDEIKPOAAALDLOAQFLTNGTARVLVAG 106
DB 135 -GVGNG-----LPALMLDADRELIVLELSF--OLETTSLOQAAATVLM 179
OY 107 HTDERGSEYNSLGERRAVAVRNYLTK-----GINQASVEIISFGEERPIAFGTNEEAM 162
DB 180 VTEDHMDR-YPFGLQYRAAKLRYEKAKVCVANNADALMPYRGADDERCVSGVMGDI 238
OY 163 SONRR 167
DB 239 HLNRQ 243

RESULT 20

US-09-828-423-5
Sequence 5, Application US/09828423
Patent No. US20020099178A1
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
Guegler, Karl J.
Patterson, Chandra
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
INHIBITOR HEAVY CHAIN PRECURSOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/828,423
FILING DATE: 05-Apr-2001
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/388,774
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0505 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 885 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENEBANK
CLONE: 91288563
SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-828-423-5
Query Match 8.8%; Score 76; DB 10; Length 885;
Best Local Similarity 25.5%; Pred. No. 28;
Matches 14; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 71 YFDESDDEIKPOAAALDLOAQFLTNGTARVAVAGHTDERGSEYNSLGERRA 125
DB 177 HFEIVDFEPGIGSLMDAESFTINDLSALTKSFSGKGVHVSFRSLDQNS 231

RESULT 21

US-10-156-761-13516
Sequence 13516, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13516
LENGTH: 135
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13516

Query Match 8.8%; Score 75.5; DB 9; Length 155;
Best Local Similarity 23.3%; Pred. No. 2.6;
Matches 38; Conservative 25; Mismatches 55; Indels 45; Gaps 7;

QY 7 IAAAAAALSVLTFMTGCAKSTQVAVNAPNPTGYGVYIGVAPLVNDDETVALKSL 66
DB 9 VAGAAVALLAL---TGCGSSGSDGAKAKAPATATG-----SLEDIAAEV 51
OY 67 PSLYVDESDDEIKPOAAALDLOAQFLTNGTARVAVAGHTDERGSEYNSLGERRAV 126
DB 52 -----KKRPDIQTDADEIRQAVCKSTRKFIATPATRGQREW---LNEAKDY 97
QY 127 AVRNLLKGINQASVEIISFGEERPIAF-----GTNEEAMS 163

Db 98 G-GEVLV-----AKWVAVGDQKVKTLRGQLGTMECTS 132

RESULT 22

US-09-815-242-5697

; Sequence 5697, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5697

; LENGTH: 794

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-5697

Query Match 8.8%; Score 75.5; DB 10; Length 794;

Best Local Similarity 23.8%; Pred. No. 27;

Matches 38; Conservative 24; Mismatches 59; Indels 39; Gaps 6;

QY 9 AAAALSVLFMTGCA--NKSTQVAVAPN--APTGY-----TG 43

Db 439 ALVASISVLVACPCALGLATPTISIMVGTGAENGILFKGGEVERTHQIDTIVLDK 498

QY 44 VIYTGAPLVD---NDETAKALASKPLSVYFDESDKIQAAALID--EQAQFLTTNQ 98

Db 499 TITNGRPVYVDYHGDDQTLQLATA-----EKDSEHPLAEALIVNAKKEQLTLTET 549

QY 99 TARVLVAGHTDEGRSREYNMSLGERRAVAVRNLYLKGKGIN 138

Db 550 TTEKAVPGHGIEATIDHHILVGNKRLMADNDISLPKHIS 589

RESULT 23

US-09-815-242-12668

; Sequence 12668, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12668

; LENGTH: 802

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12668

Query Match 8.8%; Score 75.5; DB 10; Length 802;

Best Local Similarity 23.8%; Pred. No. 27;

Matches 38; Conservative 24; Mismatches 59; Indels 39; Gaps 6;

QY 9 AAAALSVLFMTGCA--NKSTQVAVAPN--APTGY-----TG 43

Db 439 ALVASISVLVACPCALGLATPTISIMVGTGAENGILFKGGEVERTHQIDTIVLDK 498

QY 44 VIYTGAPLVD---NDETAKALASKPLSVYFDESDKIQAAALID--EQAQFLTTNQ 98

Db 499 TITNGRPVYVDYHGDDQTLQLATA-----EKDSEHPLAEALIVNAKKEQLTLTET 549

QY 99 TARVLVAGHTDEGRSREYNMSLGERRAVAVRNLYLKGKGIN 138

Db 550 TTEKAVPGHGIEATIDHHILVGNKRLMADNDISLPKHIS 589

RESULT 24

US-10-156-761-10510

; Sequence 10510, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 10510

; LENGTH: 1330

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-10510

Query Match 8.7%; Score 74.5; DB 9; Length 1330;

Best Local Similarity 25.0%; Pred. No. 72;

QY 116 YNMSIGERRAVAVRNVLGKINOASVEIISFGE 149
 Db 1136 H-----KATVYNNKKKGKGTNSDSEELPAGE 1162

RESULT 30
 US-09-815-242-5568
 ; Sequence 5568, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5568
 ; LENGTH: 1491
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-5568

Query Match 8.5%; Score 73; DB 10; Length 1491;
 Best Local Similarity 25.6%; Pred. No. 1.2e+02;
 Matches 32; Conservative 19; Mismatches 50; Indels 24; Gaps 5;

QY 31 VWAPNAPPTGYTGVITGVAPLVD-----NDETVKALASKLPISLYF-----DFDSD 77
 Db 375 VVDVPSNVAFKGALPGLLVDFKQNKVIEENDLKAGALGELPYKAMIDNHKVDDE 434
 QY 78 EIKPQAAALIDE-----QAQFLTTNQT-----ARVLVAGHTDERSGREYNSLG--ERRAV 126
 Db 435 NIQYODSQMKDETLFKLQROFAYTKKEIHKYIOELVEGKKDPDAGMGYDAPIAVLNERPE 494
 QY 127 AVRN 131
 Db 495 SLFNY 499

RESULT 31
 US-09-815-242-12162
 ; Sequence 12162, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12162
 ; LENGTH: 1502
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-09-815-242-12162

Query Match 8.5%; Score 73; DB 10; Length 1502;
 Best Local Similarity 25.6%; Pred. No. 1.2e+02;
 Matches 32; Conservative 19; Mismatches 50; Indels 24; Gaps 5;

QY 31 VWAPNAPPTGYTGVITGVAPLVD-----NDETVKALASKLPISLYF-----DFDSD 77
 Db 385 VVDVPSNVAFKGALPGLLVDFKQNKVIEENDLKAGALGELPYKAMIDNHKVDDE 444
 QY 78 EIKPQAAALIDE-----QAQFLTTNQT-----ARVLVAGHTDERSGREYNSLG--ERRAV 126
 Db 445 NIQYODSQMKDETLFKLQROFAYTKKEIHKYIOELVEGKKDPDAGMGYDAPIAVLNERPE 504
 QY 127 AVRN 131
 Db 505 SLFNY 509

RESULT 32
 US-09-738-626-4444
 ; Sequence 4444, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIRO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAOKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; PRIOR FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4444
LENGTH: 1520
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4444

Query Match
Best Local Similarity 24.6%; Pred. No. 1.3e+02;
Matches 28; Conservative 25; Mismatches 47; Indels 14; Gaps 5;

QY 53 VDNDETVKALASKLPSLYVYFDFDSEIKPOAAIIDEQOAFLLTNOTA----RVLVAGH- 107
DB 311 VNSRRSAERLSTSLNEIWMNEHDPESLSPQ---LRDPQAIMSSADVAGKAPQVYARAH 367
QY 108 ----TDERGSRREYMSLGERRAAVAVNNYLLGKINQASVEIISFGGERP-IAFG 156
DB 368 GSVSKDERATETMTLKEGRRLAV-ISTSLLEIGDMGAVDLVYQVSPSVASG 420

RESULT 33

US-10-156-761-9933
Sequence 9933, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9933
LENGTH: 851
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9933

Query Match
Best Local Similarity 21.1%; Pred. No. 62;
Matches 45; Conservative 19; Mismatches 60; Indels 89; Gaps 8;

QY 6 QIAAANAALSVLFEMGCKANKSTSOV-----MVAAPNATGYT 42
DB 483 QAAALRAESLISFLAGSVLRGETSLDALLERYETFGMESVALLERESDVAP-----WT 537
QY 43 GVIYTGAVPLVD-----NDETVALASK-LPSLYVDFDSE----IKPOAA 85
DB 538 CAGSVGYRPLVDREDAVDMPYGDHALALASGRVLP-----SDRYLAFAQAANV 590
QY 86 ILDEQOAFLLTNOTARVLAVAGHTDERGSRREYMSLGERRAAVAVNNYLLGKINQASVEI 145
DB 591 VLDRORL-----QREDAQRTLAEGNSITALLAAV 621
QY 146 SFGGERPIA-----FGTNEAMSONRRRAEL 170
DB 622 SHDLRTPLAGIKAAVSSLRSDVAVMSSEDRRAEL 654

RESULT 34
US-09-901-106-2
Sequence 2, Application US/09901106
Patent No. US20020151067A1
GENERAL INFORMATION:

APPLICANT: Garoff, Henrik
Iljeström, Peter
TITLE OF INVENTION: DNA Expression Systems Based on
Alphaviruses

NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/901,106
FILING DATE: 10-Jul-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/920,281C
FILING DATE: 13-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-901-106-2

Query Match
Best Local Similarity 23.0%; Pred. No. 2.8e+02;
Matches 26; Conservative 20; Mismatches 36; Indels 31; Gaps 5;

QY 24 ANKSTQYVAVAPNAPMTGYTGVIYTGAVPLVNDETVALASKLPSLYVDFDSEIKPOA 83
DB 1756 ASRAERVPAPARKET-----PAERT-----AFRNKLP-LTFGDFDEHYVALA 1798
QY 84 AAI-----LDEQOAFLLTNOTARVLAVAGHTDERGSRREYMSLGERRAV 126
DB 1799 SGITFGDFPDVLRGRAGAVYFSSDTG-----SGHLQKSVQHNILCAQDLAV 1847

RESULT 35
US-10-156-761-13680
Sequence 13680, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697

;; PRIOR FILING DATE: 2001-08-02
;; NUMBER OF SEQ ID NOS: 15109
;; SEQ ID NO 13680
;; LENGTH: 441
;; TYPE: PRF
;; ORGANISM: Streptomyces avermiltillis
US-10-156-761-13680

Query Match
Best Local Similarity 24.1%; Score 71.5; DB 9; Length 441;
Matches 46; Conservative 21; Mismatches 73; Indels 51; Gaps 7;

QY 6 QIAAAAALSVLTMTGANKSTQYVAPN-----APTGYTVI-----YTGVAL 52
DB 186 EYVAGGATVAVAFVGTSCAPADVGTGPNITWAAAKRYFTGKIGIDAEAGPTETIATL 245
QY 53 VDNDEYVKAASL-----PSLVTFDESDKPOAA---ILDEQACF 93
DB 246 ADSTADPVHVASDLISQAEHDPAAAVLYTDSVELDAVEKELEPQVAAATKHIDRIVPA 305
QY 94 LTTNOTARVAVAG-----HTDERSREYNMSLGERRAVA--VRNYLGGKINQASVELI 145
DB 306 LKRGQSAIYLVDSVDEGLRVADYAGAEHLEIQTADAANAADRVN-----AGAI 354
QY 146 SFGEERPIAFG 156
DB 355 FIGMPAPVSLG 365

RESULT 36
US-10-156-761-14476
; Sequence 14476, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14476
; LENGTH: 507
; TYPE: PRF
; ORGANISM: Streptomyces avermiltillis
US-10-156-761-14476

Query Match
Best Local Similarity 23.5%; Score 71.5; DB 9; Length 507;
Matches 32; Conservative 25; Mismatches 50; Indels 29; Gaps 6;

QY 41 YTGVIITGVAPLVNDDEYVKAASL-----PSLVTFDESD-----EIKP 81
DB 245 FTGGLTGRIMAAAGTAKVLELGGKNPNIVFADADFDPAVDMALAVFLHSQVCS 304
QY 82 QAAAILDEQA---QFL--TTNCTARVAVAGHTDERSREYNMSLGERRAVAVRNYLGGK 136
DB 305 AGARLLVEDSLHDFVDEVVRRRAETRLGSPDERAOTGALISAARATVEA---YVANG 361
QY 137 INQASVELISFGEERP 152
DB 362 LDEGAV--LRGGRRP 375

RESULT 37
US-09-742-659-5
; Sequence 5, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacqueline
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2862
; TYPE: PRF
; ORGANISM: GBV-B/HCV
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (945)..(1129)
; OTHER INFORMATION: chimeric region
; NAME/KEY: SITE
; LOCATION: (1579)..(1593)
; OTHER INFORMATION: chimeric region
US-09-742-659-5

Query Match
Best Local Similarity 27.5%; Score 71.5; DB 10; Length 2862;
Matches 38; Conservative 16; Mismatches 43; Indels 41; Gaps 8;

QY 3 LHIQIAAALSVLTMTGANKSTQYVAPN-----PAPGYGVITGVAPLVNDDET 58
DB 1566 LAVGVAMAYLA-----STGCV-----VIGRYLVLSKPLPGAT-----VAPVDEEET 1609
QY 59 VKAASLPSLVTFDESDKPOAAAILDEQAOFLTTNCTARVAVAGHTDERSREYNM 118
DB 1610 VEECASFP-----LEAMVAAT-DLKSTITT-----TSPFTLEFALERT 1650
QY 119 SLGERRA--VAVRNYLGG 134
DB 1651 FLGPHRAATLAIIEYCCG 1668

RESULT 38
US-09-925-301-908
; Sequence 908, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 908
; LENGTH: 248
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-925-301-908

Query Match
Best Local Similarity 8.2%; Score 71; DB 10; Length 248;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:24 ; Search time 62 Seconds

(without alignments)

850.781 Million cell updates/sec

Title: US-09-674-779B-2
Sequence: 1 MMLHIQIAAAALSVLTFM.....IAFGTNEAASQNRRAELSY 172

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip
-O=/cgn2_1/USPFO/spool/US09674779/rn1.06072003.121614.9825/app_query.fasta.1.327
-DB=Issued_Patents.NA -OPMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi
-LIST=100 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09674779 -GCN 1.1 36 -gunat -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -NCPU=6 -ICPU=3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :

Issued_Patents.NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
3: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
4: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
5: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
6: /cgn2_6/prodata/1/ina/PCITUS.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	30.7	462	2	US-08-743-637B-178
2	264	30.7	462	6	5173294-3
3	264	30.7	867	6	5173294-1
4	264	30.7	1019	1	US-07-807-049-1
5	176.5	20.5	1059	4	US-08-476-102A-3
6	175.5	20.4	645	4	US-08-572-447C-12
7	175.5	20.4	645	4	US-09-267-747C-12
8	171.5	19.9	486	2	US-08-572-447C-10
9	171.5	19.9	486	4	US-09-267-747-10
10	171.5	19.9	681	4	US-08-572-447C-14
11	171.5	19.9	681	4	US-09-267-747-14
12	165.5	19.2	1720	1	US-08-457-997B-1

C	13	165.5	19.2	1720	3	US-08-467-722A-1	Sequence 1, Appl1
C	14	16.7	3656	4	US-09-221-017B-786	Sequence 786, App	
C	15	143	4403765	4	US-09-103-840A-2	Sequence 2, App	
C	16	143	16.6	4411529	4	US-09-103-840A-1	Sequence 1, App
C	17	142	16.5	4760	4	US-09-221-017B-875	Sequence 875, App
C	18	136.5	15.9	730	2	US-08-743-637B-11	Sequence 11, App
C	19	136.5	15.9	730	2	US-08-526-840B-11	Sequence 11, App
C	20	136.5	15.9	1007	4	US-08-836-500A-1	Sequence 13, App
C	21	136.5	15.9	1008	3	US-08-836-500A-13	Sequence 13, App
C	22	136.5	15.9	1008	4	US-08-836-500A-13	Sequence 13, App
C	23	136	15.8	1727	1	US-08-129-719-14	Sequence 14, App
C	24	136	15.8	1727	1	US-08-306-871-14	Sequence 14, App
C	25	136	15.8	1727	1	US-08-306-871-14	Sequence 14, App
C	26	87.5	10.2	4268	4	US-08-469-260A-80	Sequence 80, App
C	27	87.5	10.2	8912	4	US-08-469-260A-80	Sequence 80, App
C	28	87.5	10.2	8912	4	US-08-469-260A-80	Sequence 80, App
C	29	87.5	10.2	9143	2	US-08-469-260A-11	Sequence 32, App
C	30	87.5	10.2	9143	4	US-08-469-260A-390	Sequence 390, App
C	31	81	9.4	2895	4	US-09-556-877-171	Sequence 171, App
C	32	81	9.4	2895	4	US-09-556-877-171	Sequence 171, App
C	33	81	9.4	2934	4	US-09-556-877-183	Sequence 183, App
C	34	81	9.4	2934	4	US-09-556-877-183	Sequence 183, App
C	35	78	9.1	11958	4	US-09-134-246-8	Sequence 183, App
C	36	77.5	9.0	80161	3	US-09-134-246-8	Sequence 183, App
C	37	77.5	9.0	80161	3	US-09-134-246-8	Sequence 183, App
C	38	77	8.9	4403765	4	US-09-103-840A-2	Sequence 2, App1
C	39	74	8.6	38155	4	US-09-453-702B-79	Sequence 79, App1
C	40	74	8.6	48908	4	US-09-453-702B-137	Sequence 137, App
C	41	73.5	8.5	1242	2	US-08-960-756-1	Sequence 137, App
C	42	73.5	8.5	1581	2	US-08-762-106-6	Sequence 137, App
C	43	73.5	8.5	1581	2	US-09-320-774-6	Sequence 6, App1
C	44	73.5	8.5	5849	4	US-09-134-246-6	Sequence 6, App1
C	45	73.5	8.5	5849	4	US-09-134-246-6	Sequence 6, App1
C	46	73	8.5	6002	4	US-09-495-797-44	Sequence 44, App1
C	47	73	8.5	6201	4	US-08-790-912-1	Sequence 4, App1
C	48	73	8.5	6201	4	US-08-790-912-1	Sequence 4, App1
C	49	72.5	8.4	11517	1	US-09-453-702B-252	Sequence 252, App
C	50	72.5	8.4	11517	1	US-07-920-281C-1	Sequence 1, App1
C	51	72.5	8.4	11517	1	US-08-466-277-1	Sequence 1, App1
C	52	71.5	8.4	12492	6	US-08-466-277-1	Patent No. 5206163
C	53	71.5	8.3	1108	1	US-08-181-271A-13	Sequence 13, App1
C	54	71.5	8.3	1108	1	US-08-449-315-13	Sequence 13, App1
C	55	71.5	8.3	1108	1	US-08-449-803-13	Sequence 13, App1
C	56	71.5	8.3	1108	1	US-08-449-803-13	Sequence 13, App1
C	57	71.5	8.3	1108	1	US-08-456-265A-13	Sequence 13, App1
C	58	71.5	8.3	1108	1	US-08-456-265A-13	Sequence 13, App1
C	59	71.5	8.3	1108	1	US-08-455-416-13	Sequence 13, App1
C	60	71.5	8.3	1108	1	US-08-455-244-13	Sequence 13, App1
C	61	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, App1
C	62	71.5	8.3	1108	2	US-08-457-364-13	Sequence 13, App1
C	63	71.5	8.3	1108	2	US-08-456-240-13	Sequence 13, App1
C	64	71.5	8.3	1108	2	US-08-456-240-13	Sequence 13, App1
C	65	71.5	8.3	1108	2	US-08-455-736-13	Sequence 13, App1
C	66	71.5	8.3	1108	2	US-08-971-217-13	Sequence 13, App1
C	67	71.5	8.3	1108	4	US-09-350-600-13	Sequence 13, App1
C	68	71.5	8.3	1195	1	US-08-181-271A-14	Sequence 14, App1
C	69	71.5	8.3	1195	1	US-08-444-803-14	Sequence 14, App1
C	70	71.5	8.3	1195	1	US-08-444-803-14	Sequence 14, App1
C	71	71.5	8.3	1195	1	US-08-445-043-14	Sequence 14, App1
C	72	71.5	8.3	1195	1	US-08-445-265A-14	Sequence 14, App1
C	73	71.5	8.3	1195	1	US-08-455-416-14	Sequence 14, App1
C	74	71.5	8.3	1195	1	US-08-455-244-14	Sequence 14, App1
C	75	71.5	8.3	1195	2	US-08-457-364-14	Sequence 14, App1
C	76	71.5	8.3	1195	2	US-08-456-262-14	Sequence 14, App1
C	77	71.5	8.3	1195	2	US-08-456-262-14	Sequence 14, App1
C	78	71.5	8.3	1195	2	US-08-456-240-14	Sequence 14, App1
C	79	71.5	8.3	1195	2	US-08-455-736-14	Sequence 14, App1
C	80	71.5	8.3	1195	4	US-08-971-217-14	Sequence 14, App1
C	81	71.5	8.3	4411529	4	US-09-350-600-14	Sequence 14, App1
C	82	71	8.2	1765	1	US-09-103-840A-1	Sequence 1, App1
C	83	71	8.2	1765	2	US-08-459-586-15	Sequence 15, App1
C	84	71	8.2	18994	1	US-08-282-696-4	Sequence 4, App1
C	85	69	8.0	2797	2	US-08-282-696-4	Sequence 4, App1
C	86	69	8.0	2797	2	US-08-954-333-8	Sequence 8, App1

86 69 8.0 2820 5 PCT-US94-05905-19 Sequence 19, Appl
87 69 8.0 5117 3 US-08-854-585-1 Sequence 1, Appl
88 69 8.0 5117 5 PCT-US95-05512-1 Sequence 1, Appl
89 68.5 8.0 531 2 US-08-809-267-9 Sequence 9, Appl
90 68.5 8.0 531 2 PCT-US95-13662A-9 Sequence 3, Appl
91 68.5 8.0 1359 1 US-07-618-312A-3 Sequence 7, Appl
92 68.5 8.0 1359 1 US-08-110-786A-7 Sequence 6, Appl
93 68.5 8.0 1359 1 US-08-280-228-3 Sequence 39, Appl
94 68.5 8.0 1858 1 US-08-668-381A-6 Sequence 57, Appl
95 68.5 8.0 25165 4 US-09-453-702B-39 Sequence 60, Appl
96 68 7.9 955 3 US-08-784-582-60 Sequence 72, Appl
97 68 7.9 2356 3 US-08-784-582-72 Sequence 314, Appl
98 68 7.9 3190 4 US-09-221-017B-314 Sequence 242, App
99 68 7.9 31880 4 US-09-453-702B-242

ALIGNMENTS

RESULT 1
US-08-743-637B-178
Sequence 178, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESSES:
ADDRESSEE: OUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5500
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 178:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Haemophilus influenzae
US-08-743-637B-178
Alignment Scores: 9,21e-27 Length: 462
Pred. No.: 264.00 Matches: 51
Score: 264.00

Percent Similarity: 68.93%
Best Local Similarity: 49.51%
Query Match: 30.66%
DB: 2
Gaps: 0

US-09-674-779B-2 (1-172) x US-08-743-637B-178 (1-462)

70 ValTyrPheAspPheAspSerAspGluIleTyrProGlnAlaAlaIleLeuAspGlu 89
151 GGTATATTGGTTTGTGATTAATACGACATCCGCGTCAATCTTACATCG 210
90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaValLeuValAlaGlyHisThrAsp 109
211 CACGACGATATTATTAATGCAACGCCAGCTGCTAAAGTATGTAAGTAATCTGAT 270
110 GlnArgGlySerArgGlyIleThrAsnMetSerLeuGlyGlnArgAlaValAlaValArg 129
271 GAACGTGTACACCAATACCAATACCAATACCAATACCAATACCAATACCAATACCA 330
130 AsnTyrLeuLeuGlyIleAsnGlnAlaSerValGluIleLeuSerPheGlyGlu 149
331 GGTATATTACGAGTAAGGTGTGATGCTGTAAATTAGGCACAGTATCTTACGGTGA 390
150 GlnArgProIleAlaPheGlyThrAsnGlnAlaIleTyrSerGlnAsnArgAlaGlu 169
391 GAAACCTCAGTATTAGTGTACGATGAGAGCTGCAATCTTAAACCGTGTGACAGT 450
170 LeuSerTyr 172
451 TTACGCTAC 459

RESULT 2

5173294-3
Patent No. 5173294
APPLICANT: MURPHY, TIMOTHY F.; APICELLA, MICHAEL A.
TITLE OF INVENTION: DNA PROBE FOR THE IDENTIFICATION
OF HAEMOPHILUS INFLUENZAE
NUMBER OF SEQUENCES: 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/92,948
FILING DATE: 08-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 932,872
FILING DATE: 18-NOV-1986
SEQ ID NO: 3
LENGTH: 462
5173294-3

Alignment Scores:

Pred. No.: 9,21e-27 Length: 462
Score: 264.00 Matches: 51
Percent Similarity: 68.93% Conservative: 20
Best Local Similarity: 49.51% Mismatches: 32
Query Match: 30.66% Indels: 0
DB: 6 Gaps: 0

US-09-674-779B-2 (1-172) x 5173294-3 (1-462)

70 ValTyrPheAspPheAspSerAspGluIleTyrProGlnAlaAlaIleLeuAspGlu 89
151 GGTATATTGGTTTGTGATTAATACGACATCCGCGTCAATCTTACATCG 210
90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaValLeuValAlaGlyHisThrAsp 109
211 CACGACGATATTATTAATGCAACGCCAGCTGCTAAAGTATGTAAGTAATCTGAT 270
110 GlnArgGlySerArgGlyIleThrAsnMetSerLeuGlyGlnArgAlaValAlaValArg 129
271 GAACGTGTACACCAATACCAATACCAATACCAATACCAATACCAATACCAATACCA 330
130 AsnTyrLeuLeuGlyIleAsnGlnAlaSerValGluIleLeuSerPheGlyGlu 149
331 GGTATATTACGAGTAAGGTGTGATGCTGTAAATTAGGCACAGTATCTTACGGTGA 390

QY 150 GUARGPRLIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgAlaGlu 169
 |||:|||||
 Db 391 GAAAAACCTGCAGATTAGTACGATGAAAGCTGCAATATTCTAAACCGTCTGCAGTG 450
 |||:|||||
 QY 170 LeuSerTyr 172
 |||:|||||
 Db 451 TTAGCGTAC 459
 |||:|||||

RESULT 3
 5173294-1
 Patent No. 5173294
 APPLICANT: MURPHY, TIMOTHY F., APICELLA, MICHAEL A.
 TITLE OF INVENTION: DNA PROBE FOR THE IDENTIFICATION
 OF HAEMOPHILUS INFLUENZAE
 NUMBER OF SEQUENCES: 3
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/92,948
 FILING DATE: 08-OCT-1987
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 932,872
 FILING DATE: 18-NOV-1986
 SEQ ID NO: 1:
 LENGTH: 867
 5173294-1

Alignment Scores:
 Pred. No.: 2.38e-26 Length: 867
 Score: 264.00 Matches: 51
 Percent Similarity: 68.93% Conservative: 20
 Best Local Similarity: 49.51% Mismatches: 32
 Query Match: 30.66% Indels: 0
 DB: 6 Gaps: 0

US-09-674-779b-2 (1-172) x 5173294-1 (1-867)

QY 70 ValTyrPheAspPheSerAspGluIleYsProGlnAlaAlaIleLeuAspGlu 89
 |||:|||||
 Db 218 GTATATTGGTTTATTAATACGATCACCAGTGAATACCTTCAATCTTGATCGG 277
 |||:|||||
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValIleValAlaGlyHisThrAsp 109
 |||:|||||
 Db 278 CACGACGATATTTAATGCAACGCCAGCTGCTAAGATTAGTGAAGGAAATACGAT 337
 |||:|||||
 QY 110 GluArgGlySerArgGlnIleTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 |||:|||||
 Db 338 GAACGCTGTACACCGAATACAAATCGCATTAGAGCAACGTCGTCAGATGAGTTAA 397
 |||:|||||
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 |||:|||||
 Db 398 GGTATTATTAGCAGTAAGGTGTGATGCTGTAATATTAGCAGATCTTACGTTGA 457
 |||:|||||
 QY 150 GluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgAlaGlu 169
 |||:|||||
 Db 458 GAAAAACCTGCAGATTAGTACGATGAAAGCTGCAATATTCTAAACCGTCTGCAGTG 517
 |||:|||||

QY 170 LeuSerTyr 172
 |||:|||||
 Db 518 TTAGCGTAC 526
 |||:|||||

RESULT 4
 US-07-807-049-1
 Sequence 1, Application US/07807049
 Patent No. 5300632
 GENERAL INFORMATION:
 APPLICANT: Murphy, Timothy F
 APPLICANT: Apicella, Michael A
 TITLE OF INVENTION: A Method for Purifying an Outer Membrane
 PROTEIN OF HAEMOPHILUS INFLUENZAE
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Alan S. Korman, Esq.
 STREET: 1600 Empire Tower

CITY: Buffalo
 STATE: NY
 COUNTRY: U.S.A.
 ZIP: 14202

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/807,049
 FILING DATE: 19911212
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/330,229
 FILING DATE: 29-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/092,948
 FILING DATE: 08-OCT-1987

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/932,872
 FILING DATE: 18-NOV-1986
 ATTORNEY/AGENT INFORMATION:
 NAME: KORMAN ESQ., ALAN S
 REGISTRATION NUMBER: 32,932

REFERENCE/DOCKET NUMBER: 19226/00028
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 853-8100
 TELEFAX: (716) 853-8109
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1019 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Hemophilus influenzae
 STRAIN: 1479
 INDIVIDUAL ISOLATE: 1479
 HAPLOTYPE: N/A
 IMMEDIATE SOURCE:
 LIBRARY: N/A
 CLONE: N/A

POSITION IN GENOME:
 MAP POSITION: Unknown
 PUBLICATION INFORMATION:
 AUTHORS: Nelson, M B
 AUTHORS: Apicella, M A
 AUTHORS: Murphy, T F
 AUTHORS: Vankulien, H
 AUTHORS: Spetella, L D
 AUTHORS: Rekosh, D M

TITLE: Molecular Analysis of P6: The cloning and
 TITLE: sequencing of an Outer Membrane Protein
 JOURNAL: Infect. Immun.
 VOLUME: 56
 PAGES: 128-134

DATE: 1988
 PUBLICATION INFORMATION:
 DOCUMENT NUMBER: US 4,427,782
 FILING DATE: 24-JAN-1984
 PUBLICATION INFORMATION:
 DOCUMENT NUMBER: US 4,474,758
 FILING DATE: 02-OCT-1984

PUBLICATION INFORMATION:
 DOCUMENT NUMBER: US J.CLIN.INVEST.V.78
 PUBLICATION DATE: 01-OCT-1986
 RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1020 TO 1027

[illegible]

Alignment Scores:	
Pred. No.:	1.41e-14
Score:	175.50
Percent Similarity:	46.58%
Length:	64
Matches:	49
Conservative:	19

Best Local Similarity: 33.56% Mismatches: 68
 Query Match: 20.38% Indels: 11
 DB: 4 Gaps: 2

US-09-674-779b-2 (1-172) x US-09-267-747-12 (1-645)

QY 37 AlapProthrglyTyrThrglyValIleTyrThrglyAlaProLeuValAspAsnasp 56
 |||||
 DB 26 GCTCCGACTCCGACACGACGCGCTGTGACACAC-GTGCAGACAGTGCAGACACCGC 84
 QY 57 GluThrValIleAlaLeuAlaSerIleuProSerLeu----- 69
 |||||
 DB 85 GCCACGTGACCGCTTACCGCCGCTCCGCGCTCCGCGGAGTGTACGCTACG 144
 QY 70 -----ValTyrPheAspPheAspSerAspGluIleTyrProGluAlaIleIleu 87
 |||||
 DB 145 CTGGAGCTGAGTGTGACTTGCAGACGTCACAGGTCAAGAGACAGACAGCTACGCTACATC 204
 QY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValaGlyHis 107
 |||||
 DB 205 AAGACCTGGCCGACCTTCATGAGCAGTACCGCTCCACTCCACACCGCTGAGGTCTAT 264
 QY 108 ThrAspGluArgIleSerArgGluTyrAsnMetSerLeuGlyGluArgArgAlaIle 127
 |||||
 DB 265 ACCGACTCCGCTCGTACCGACGCTTACACCAAGCTGTCCGAGGTGTCCACACGCC 324
 QY 128 ValArgAspTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleIleSer 146
 |||||
 DB 325 GTTCGTGACGTACTGTGTCACAGGACGAGTGTGAGGTGTGCGCTGAAACGCTGTGCGT 384
 QY 147 PheGlyGluArgProIleAlaPheGlyThrAsnGluGlnAlaTrpSerGlnAsnArg 166
 |||||
 DB 385 TACGGGACGACCGCCGCGGTGGCCGACACGACCGACGCTGAGGCGCGCTATCAACCGT 444
 QY 167 ArgAlaGluLeuSerTyr 172
 |||||
 DB 445 CGCGTTGAAGACGCCAC 462

RESULT 8

US-08-572-447C-10
 ; Sequence 10, Application US/08572447C
 ; Patent No. 5955090

GENERAL INFORMATION:
 ; APPLICANT: Knapp, Bernhard
 ; APPLICANT: Hungerer, Klaus-Dieter
 ; APPLICANT: Broker, Michael
 ; APPLICANT: Von Specht, Bernd-Ulrich
 ; APPLICANT: Domdey, Horst
 ; TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-Oprf
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
 ; ADDRESS: Dunnet
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/572.447C
 ; FILING DATE: 14-DEC-1995
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: EP 94120023.0
 ; FILING DATE: 16-DEC-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: M. Paul Barker

REGISTRATION NUMBER: 32,013
 ; REFERENCE/DOCKET NUMBER: 05552.1395-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 408-4000
 ; TELEFAX: (202) 408-4400
 ; INFORMATION FOR SRO ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 486 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Pseudomonas aeruginosa
 ; TISSUE TYPE: Serotype 6; ATCC 33354
 ; FEATURE:

NAME/KEY: CDS
 ; LOCATION: 1..483
 ; OTHER INFORMATION: /note= "Sequence is coding for oprf"
 ; Patent No. 5955090
 ; OTHER INFORMATION: C-terminus"

Alignment Scores:
 ; Pred. No.: 3,19e-14 Length: 486
 ; Score: 171.50 Matches: 48
 ; Percent Similarity: 46.15% Conservative: 18
 ; Best Local Similarity: 33.57% Mismatches: 67
 ; Query Match: 19.92% Indels: 11
 ; DB: 2 Gaps: 2

US-09-674-779b-2 (1-172) x US-08-572-447C-10 (1-486)

QY 37 AlapProthrglyTyrThrglyValIleTyrThrglyAlaProLeuValAspAsnasp 56
 |||||
 DB 32 GCTCCGACTCCGACACGACGCGCTGTGACACAC-GTGCAGACAGTGCAGACACCGC 90
 QY 57 GluThrValIleAlaLeuAlaSerIleuProSerLeu----- 69
 |||||
 DB 91 GCCACGTGACCGCTTACCGCCGCTCCGCGCTCCGCGGAGTGTACGCTACG 150
 QY 70 -----ValTyrPheAspPheAspSerAspGluIleTyrProGluAlaIleIleu 87
 |||||
 DB 151 CTGGAGCTGAGTGTGACTTGCAGACGTCACAGGTCAAGAGACAGACAGCTACGTCATC 210
 QY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValaGlyHis 107
 |||||
 DB 211 AAGACCTGGCCGACCTTCATGAGCAGTACCGCTCCACTCCACACCGCTGAGGTCTAT 270
 QY 108 ThrAspGluArgIleSerArgGluTyrAsnMetSerLeuGlyGluArgArgAlaIle 127
 |||||
 DB 271 ACCGACTCCGCTCGTACCGACGCTTACCAACGAGAGTGTCCGAGCTGTGACACGCC 330
 QY 128 ValArgAspTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleIleSer 146
 |||||
 DB 331 GTTCGTGACGTACTGTGTCACAGCAGTACGCTGTGAGAGTGTGCGGTGAAACGCTGTGCGT 390
 QY 147 PheGlyGluArgProIleAlaPheGlyThrAsnGluGlnAlaTrpSerGlnAsnArg 166
 |||||
 DB 391 TACGGGACGTCCCGCGGTGCGACACGACGACCGCTGAGGCGCGGTATCAACCGT 450
 QY 167 ArgAlaGlu 169
 |||||
 DB 451 CGCGTTGAA 459

RESULT 9

US-09-267-747-10
 ; Sequence 10, Application US/09267747
 ; Patent No. 6300102
 ; GENERAL INFORMATION:
 ; APPLICANT: Knapp, Bernhard

APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Dornay, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-Oprf
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/267,747
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/572,447
FILING DATE: 14-DEC-1995
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
FAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: YES
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
TISSUE TYPE: Serotype 6; ATCC 33354
FEATURE:
NAME/KEY: CDS
LOCATION: 1..483
OTHER INFORMATION: /note="Sequence is coding for oprf
Patent No. 6300102
OTHER INFORMATION: C-terminus"
US-09-267-747-10
Alignment Scores:
Pred. No.: 3.19e-14 Length: 486
Score: 171.50 Matches: 48
Percent Similarity: 46.15% Conservative: 18
Best Local Similarity: 33.57% Mismatches: 67
Query Match: 19.92% Indels: 11
DB: 4 Gaps: 2
US-09-674-779b-2 (1-172) x US-09-267-747-10 (1-486)
QY 37 AlaprotHrGlyTyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnsp 56
Db 32 GCTCCGACTCCGACGACGAGCGGCTGCGACAAC-GTGCACAAGTCCCGGACACCCCG 90
QY 57 GluThValIleAlaLeuAlaSerIleuProSerLeu----- 69
Db 91 GCCAAGCTACCGCTTACGCCAAGCGCTCCCGGCTGTGCTGCCAAGTCTACGCTACAG 150

QY 70 -----ValTyrPheAspPheAspSerAspGluIleIleProGlnAlaAlaIleLeu 87
Db 151 CTGACGCTGAACTTGCACCTGACAGTCCAAAGTCAAGACGCTACGCTGACATC 210
QY 88 AspGluGlnAlaGlnPheLeuThrThrasnGlnThrAlaArValIleValAlaGlyHis 107
Db 211 AGAAGCTGCCGACTTATGATGAGCAGACCGCTTCCACTTCCACCGCTTGAAGGTCAT 270
QY 108 ThrAspGluArGlySerArGlyIuTyrAsnMetSerLeuGlyGluArGAlaValAla 127
Db 271 ACCGACTCCGCTCGTACGACGCTTACCAACGAGGCTGCTCCGAGCGCTGCAACGCC 330
QY 128 ValArgAsnTyrLeuLeuGlyLys--GlyIleAsnGlnAlaSerValGluIleIleSer 146
Db 331 GTTCGTGACGACTGTGTACGAGTGTGAGAGTGTGCGTACACGCTGCGT 390
QY 147 PheGlyGluGluArGProIleAlaPheGlyThrAsnGluGlnAlaTyrSerGlnAsnArg 166
Db 391 TACGCGAGTCCCGCGCGGTTGCGACAAACGCAACCGCTGAGAGCGCGCTATACACCT 450
QY 167 ArgAlaGlu 169
Db 451 CGCGTTGAA 459
RESULT 10
US-08-572-447C-14
Sequence 14, Application US/08572447C
Patent No. 595050
GENERAL INFORMATION:
APPLICANT: Knapp, Bernhard
APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Broker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Dornay, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-Oprf
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/572,447C
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
FAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 681 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO

ANTI-SENSE: YES
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
TISSUE TYPE: Serotype 6; ATCC 33354
FEATURE:
NAME/KEY: CDS
LOCATION: 1..678
OTHER INFORMATION: /note="Sequence is coding for oprI
Patent No. 5955090
OTHER INFORMATION: without signal sequence and oprf C-terminus"
US-08-572-447C-14

Alignment Scores:
Pred. No.: 5,31e-14 Length: 681
Score: 171.50 Matches: 48
Percent Similarity: 46.15% Conservative: 18
Best Local Similarity: 33.57% Mismatches: 67
Query Match: 19.92% Indels: 11
DB: 2 Gaps: 2

US-09-674-779B-2 (1-172) x US-08-572-447C-14 (1-681)

QY 37 AlaProthrlgYrThrlgYValIleYrThrlgYValAlaProleuValaAspasp 56
|||||
DB 227 GCTCGACTCCGACACGACGCGCTGCGACAAAC-GTCGACAAAGTCCCGGACACCCG 285
57 GluThrVallyAlaLeuAlaSerlysluProserleu----- 69
DB 286 GCCAAGCTGACCGTTGACGCCAAGCGCTGCGCGGCGAAGTGTACGCGTACG 345
70 -----ValTyrPheaspPheaspSeraspGluIlelyProGlnAlaAlaIleleu 87
DB 346 CTGACGCTGAGTTCGACTTCGACAGTCAAGTCAAGAGAGACAGTACGCTGACATC 405
QY 88 AspGluGlnAlaGlnPheleuThrThrasnGlnThraAlaGValleuValaIlelyHis 107
DB 406 AAGAACTGCGCCGACTTCATAGACAGTACCCGCTCCACTCCACCCTTGAAGTCAT 465
QY 108 ThrAspGluArglySerArgGluTyrAsnMetSerleuGlyGluArgAlaValAla 127
DB 466 ACCGACTCCGTCGTCGACGCGTTCACACGAGAGTGTCCGAGCGCTGCCAACGCC 525
QY 128 ValArgAsnTyrLeuLeuGlyLys---GlylleAsnGlnAlaSerValGluIleIleSer 146
DB 526 GTTCGTGACGTCAGTGTGTCACAGTACGTCGTGAGAGTGTCCGTCGACGTCGCT 585
QY 147 PheGlyGluGlnArgProIleAlaPheGlyThrasnGlnGluAlaIlePserGlnAsnArg 166
DB 586 TACGCGAGTCCCGCGCTGTCGCGACAGCCACCGCTGAGAGCGCGCTATCACCGCT 645
QY 167 ArgAlaGln 169
DB 646 CGCGTTGAA 654

RESULT 11
US-09-267-747-14
Sequence 14, Application US/09267747
Patent No. 6300102
GENERAL INFORMATION:
APPLICANT: Knapp, Bernhard
APPLICANT: Hungerer, Klaus-Dieter
APPLICANT: Brocker, Michael
APPLICANT: Von Specht, Bernd-Ulrich
APPLICANT: Domdey, Horst
TITLE OF INVENTION: Immunogenic Hybrid Protein Oprf-OprI
TITLE OF INVENTION: Derived From Pseudomonas aeruginosa Membrane Proteins.
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/267,747
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/572,447
FILING DATE: 14-DEC-1995
APPLICATION NUMBER: EP 94120023.0
FILING DATE: 16-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: M. Paul Barker
REGISTRATION NUMBER: 32,013
REFERENCE/DOCKET NUMBER: 05552.1395-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 681 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: YES
ORIGINAL SOURCE:
ORGANISM: Pseudomonas aeruginosa
TISSUE TYPE: Serotype 6; ATCC 33354
FEATURE:
NAME/KEY: CDS
LOCATION: 1..678
OTHER INFORMATION: /note="Sequence is coding for oprI
Patent No. 6300102
OTHER INFORMATION: without signal sequence and oprf C-terminus"
US-09-267-747-14

Alignment Scores:
Pred. No.: 5,31e-14 Length: 681
Score: 171.50 Matches: 48
Percent Similarity: 46.15% Conservative: 18
Best Local Similarity: 33.57% Mismatches: 67
Query Match: 19.92% Indels: 11
DB: 4 Gaps: 2

US-09-674-779B-2 (1-172) x US-09-267-747-14 (1-681)

QY 37 AlaProthrlgYrThrlgYValIleYrThrlgYValAlaProleuValaAspasp 56
|||||
DB 227 GCTCGACTCCGACACGACGCGCTGCGACAAAC-GTCGACAAAGTCCCGGACACCCG 285
QY 57 GluThrVallyAlaLeuAlaSerlysluProserleu----- 69
DB 286 GCCAAGCTGACCGTTGACGCCAAGCGCTGCGCGGCGAAGTGTACGCGTACG 345
QY 70 -----ValTyrPheaspPheaspSeraspGluIlelyProGlnAlaAlaIleleu 87
DB 346 CTGACGCTGAGTTCGACTTCGACAGTCAAGTCAAGAGAGACAGTACGCTGACATC 405
QY 88 AspGluGlnAlaGlnPheleuThrThrasnGlnThraAlaGValleuValaIlelyHis 107
DB 406 AAGAACTGCGCCGACTTCATAGACAGTACCCGCTCCACTCCACCCTTGAAGTCAT 465
QY 108 ThrAspGluArglySerArgGluTyrAsnMetSerleuGlyGluArgAlaValAla 127
DB 466 ACCGACTCCGTCGTCGACGCGTTCACACGAGAGTGTCCGAGCGCTGCCAACGCC 525

QY 35 ProAsnAlaProThrGlyTyr-----ThrGlyValIleYr 46
 Db 1001 CCAATACCGCAATTACTACAACCCTTGGATTGTTGATCAATGCGGATATTTCTTAC 1060

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Gollrick, Mary E.
REGISTRATION NUMBER: 34, 829
REFERENCE/DOCKET NUMBER: 22727/00102
TELECOMMUNICATION INFORMATION:
TELEPHONE: (216) 622-8458
TELEFAX: (216) 241-0816
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1720 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: non-typable Haemophilus influenzae
STRAIN: 1128
FEATURE:
NAME/KEY: CDS
LOCATION: 407..1483

US-08-467-722A-1

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score: 1.39e-12	1720	47	26	63	21	4
Percent Similarity: 46.508						
Best Local Similarity: 29.948						
Query Match: 19.228						

US-09-674-779B-2 (1-172) x US-08-467-722A-1 (1-1720)

```

QY 35 ProAsnAlaProThrGlyTyr-----ThrGlyValIleTyr 46
DB 1001 CCAAAATACCGCAATTACTACAAACCTTGATGTTGTATCATCGCGGTAATTTCTTAC 1060
QY 47 -----ThrGlyValAlaProLeuValAspAsnAspGluThrValIleValAlaLeuAla 63
DB 1061 CGTTTCGTCAGGCGAGACACCACTTGTTCGACGACCTGAATGTGAAGCAAACTTTC 1120
QY 64 SerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAla 83
DB 1121 AGC---TTAAATTCGATGTACTTTCGCAATTTGGTAAAGCAAACTTAAACCTCAACGA 1177
QY 84 AlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeu 103
DB 1178 CAAGCTACATTAGACAGCGCTCTATGCGCAATTTCAACAAGTTAAAGTCGAAAGTACT 1237
QY 104 ValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGluArg 123
DB 1238 GTTCGCTGATACCTACACGCTATTGTTCTGACGCGTTCACGTAACCTTCTCAAGAA 1297
QY 124 ArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGlu 143
DB 1298 CGTCGACGATTCGATCTACTACTTGTCTAAAGTGTCGACGACGACGACCACTTCA 1357
QY 144 IleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu-----160
DB 1358 GCACACTGCTTACGCTGACGAACCACTGACGCGGCACTTGTACCACTTAAGT 1417
QY 161 -----AlaTyrSerGlnAsnArgArgAlaGluLeuSer 171
DB 1418 CGTAAAGCACTTATCGCTTGTCTGCTCCAGACCGCTGCTGAAGAAATTCGA 1468

RESULT 14
US-09-221-017B-786/C
; Sequence 786, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221.017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546

```

```

; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 786:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...3656

US-09-221-017B-786

Alignment Scores:
Pred. No.: 3.49e-09 Length: 3656
Score: 144.00 Matches: 48
Percent Similarity: 37.07% Conservative: 28
Best Local Similarity: 23.41% Mismatches: 83
Query Match: 16.72% Indels: 46
DB: 4 Gaps: 5

US-09-674-779B-2 (1-172) x US-09-221-017B-786 (1-3656)
QY 8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAla--AsnLys 26
DB 2438 AGTGAAGTACGCTCATGCTTTCGCGGCTCTGCTGCGCGGTCGCGACGACGACAT 2379
QY 27 SerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyr 46
DB 2378 ATGCCAAAGCGCGGCTTATCGGC-----GCCGAGTAGAGAGTGCATTGCT 2331
QY 47 ThrGlyValAlaProLeuValAspAsn-----55
DB 2330 GCCGAGTAGAGTACGTAACCGGAAATACGGCTGCGGTGCCATCGCTGACTGCAGTC 2271
QY 56 -----AspGluThrValIleValAlaLeuAla 63
DB 2270 GGTGAGCAGCGCGTCTCTCATCGAAAGAAAGATGACAGCAAGAAAAAGAACTGGAG 2211
QY 64 SerLysLeuProSer-----LeuValTyr 71
DB 2210 GCCGAGTAGAGTACGTAACCGGAAATACGGCTGCGGTGCCATCGCTGACTGCAGTC 2151
QY 72 PheAsp-----PheAspSerAspGluIleLysProGlnAlaAlaAla 85
DB 2150 TTCGATAGCGGATCTCTTTCGACGAACTCCACACTCTGAGTCCCACTTCAAGCACT 2091
QY 86 IleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAla 105
DB 2090 GCGCTGACGAGTTCGTCGCAACATGAACAAACCCCGACGCGATATTCGATCGTA 2031
QY 106 GlyHisThrAspGluArgLysSerArgGluTyrAsnMetSerLeuGlyGluArgValAla 125
DB 2030 GGCCATACGAGCAATACCGGCTCCGACAAAGATCAACGATCCCTCTGTGAGAGAGCTGCA 1971

```

QY 126 ValAlaValArgAsnTyrLeuLeuGlyGlyIleAsnGlnAlaSerValGluIle 145
Db 1970 GCCAGCGTATATCTTCTCTGAATCTCAGGGTGTGAGTATGTCCGATGCCAGCCAA 1911
QY 146 SerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsn 165
Db 1910 GGGCGTGGAGCCAGTACGAGCGGTGTGACACATACACAGTCCGCGACGTTCCGCCAAC 1851
QY 166 ArgArgAlaGluLeu 170
Db 1850 CGCGGTGTGAGGTT 1836

RESULT 15

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Alignment Scores:

Pred. No.:	0.00021	Length:	4403765
Score:	143.00	Matches:	52
Percent Similarity:	34.98%	Conservative:	26
Best Local Similarity:	23.32%	Mismatches:	85
Query Match:	16.61%	Indels:	60
DB:	4	Gaps:	4

US-09-674-779B-2 (1-172) x US-09-103-840A-2 (1-4403765)

QY 8 AlaAlaAlaAlaAlaAlaAlaSerValLeuThrPheMetThrGlyCysAlaAsnLysSer 27
Db 1003111 GCCGCCAAGGGCGCTTGTATGACGGGCTTAACGGCTTGTCTCCGGCGGTGAACGTC 1003170
QY 28 ThrSerGlnValMetValAlaPro-----AsnAlaProThr 39
Db 1003171 ATCGACCAAGATTCACGTCGATCCCGTGTGCGATCACTTGATTTCTCAAGTCCGGAACA 1003230
QY 40 GlyTyrThrGly-----Val 44
Db 1003231 GTTTTCACCGCCAGCGCGGATTCCTGATTTTGGCTCAAGTGAAGGACACACCGTC 1003290
QY 45 IleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys-----AlaLeuAla 63
Db 1003291 ACCTTGACCGGACACTGCCCTTCATCCGACGACACAGGACGAGTGAAGCGCGCGGACAC 1003350
QY 64 SerLysLeuProSerLeu----- 69
Db 1003351 AGCACTGGCTGTGATGAATAATCGTTAAACAATATTGAGTTACGGGACGACCGCCCA 1003410
QY 69 ----- 69
Db 1003411 GGAACCCCGGCGCTCGGGCCATGTGCGGACCTGCATACGCAATACGCAATAGCCGTGACGGGT 1003470
QY 70 -----ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaAlaIleLeu 87

Db 1003471 GGAACCATCGCTTGGCAGACAGCGGGCTAGTGTGATCCAGCCGACATGAATCTGC 1003530
QY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaAlaGlyValIleValAlaGlyHis 107
Db 1003531 AACCGGTATGCGGACAGCTCAAGGATGTCGAGCGTCCGGTGGAGATCAACGCTTAC 1003590
QY 108 ThrAspGluArgLysSerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAla 127
Db 1003591 ACCGACACACCGGACGAGGATGATCAATTCCTGTTGAGCGCTCAAGCCAGACATA 1003650
QY 128 ValArgAsnTyrLeuLeuGlyGlyIleAsnGlnAlaSerValGluIleIleSerPhe 147
Db 1003651 GTCCCGGACTACCTGTGTCGCCGAGATTCCCGCGCATCAATTCACCGCTGGGTC 1003710
QY 148 GlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 167
Db 1003711 GGTTCGGTGAATCCGATGCGCCAGCACGCCACCGGAGGGCGCGCAAGATCGTCC 1003770
QY 168 AlaGluLeu 170
Db 1003771 GTCGAGATC 1003779

RESULT 16

US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:

Pred. No.:	0.00021	Length:	4411529
Score:	143.00	Matches:	52
Percent Similarity:	34.98%	Conservative:	26
Best Local Similarity:	23.32%	Mismatches:	85
Query Match:	16.61%	Indels:	60
DB:	4	Gaps:	4

US-09-674-779B-2 (1-172) x US-09-103-840A-1 (1-4411529)

QY 8 AlaAlaAlaAlaAlaAlaAlaSerValLeuThrPheMetThrGlyCysAlaAsnLysSer 27
Db 1003110 GCCGCCAAGGGCGCTTGTATGACGGGCTTAACGGCTTGTCTCCGGCGGTGAACGTC 1003169
QY 28 ThrSerGlnValMetValAlaPro-----AsnAlaProThr 39
Db 1003170 ATCGACCAAGATTCACGTCGATCCCGTGTGCGATCACTTGATTTCTCAAGTCCGGAACA 1003229
QY 40 GlyTyrThrGly-----Val 44
Db 1003230 GTTTTCACCGCCAGCGCTGCCATTCCTGATTTTGGCTCAAGTGAAGGACACCGTC 1003289
QY 45 IleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys-----AlaLeuAla 63
Db 1003290 ACCTTGACCGGAACTGCCCTTCATCCGACGACGACGAGTGAAGCGCGCGGACAC 1003349
QY 64 SerLysLeuProSerLeu----- 69

```

Db 1003350 AGCACTGGCGCTGACATAAAATCTTAACAAATATTAGTTACGGGACAGCCACGCCA 1003409
||| |||
QY 69 ----- 69
Db 1003410 GGACCCCGGCGCTCCGCCCATGTGCCACCTGCAATCAACCATCATCGCTGACGGGT 1003469
QY 70 -----ValTyrPheAspSerAspGluIleLysProGlnAlaIleLeu 87
Db 1003470 GGACCCATCGCGTTTGGCAACGACAGGGGCTAGTCTGATCCAGCCGACTATGAATCCCTG 1003529
QY 88 AspGluGlnAlaGlnPheLeuThrTrpAsnGlnThrAlaArgValLeuValAlaGlyHis 107
Db 1003530 AACCGGGTACCCGACCAAGCATGCGATGTCGGACGCTCGGGTGGCATCAACAGCTTAC 1003589
QY 108 ThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyLysArgAlaValAla 127
Db 1003590 ACCGACAAACACCGCGACCGCAAGTATCAATATCCGTTAGCGCTCAGCGAGCAGATA 1003649
QY 128 ValArgAsnTyrLeuLeuGlyLysGlyTleAsnGlnAlaSerValGluIleIleSerPhe 147
Db 1003650 GTCGCCGACTACCTGCTGGCCCGGAGTTGCCGGCGATCACATTCACCGCTGGGCTCTC 1003709
QY 148 GlyGluGlnArgProIleAlaPheGlyThrAsnGluGlnAlaTrpSerGlnAsnArgArg 167
Db 1003710 GGTTCGGTGAATCCGATCCGACGACACCCACCGCAGGGGCGCGCAAGATCTGTCG 1003769
QY 168 AlaGluLeu 170
Db 1003770 GTCGAGATC 1003778

RESULT 17
US-09-221-017B-875
; Sequence 875, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, BRUCE C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: MONROY, GLADYS H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600

```

```

TELEFAX: 650-494-0792
TELEX: 706141
; INFORMATION FOR SEQ ID NO: 875:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1...4760
US-09-221-017B-875

Alignment Scores:
Pred. No.: 9.69e-09 Length: 4760
Score: 142.00 Matches: 27
Percent Similarity: 55.42% Conservative: 19
Best Local Similarity: 32.53% Mismatches: 37
Query Match: 16.49% Indels: 0
Gaps: 0

US-09-674-779b-2 (1-172) x US-09-221-017B-875 (1-4760)
QY 70 ValTyrPheAspSerAspGluIleLysProGlnAlaIleLeuAspGlu 89
Db 2100 ATTTCCTATGATTTCGATAAAGCTACTCTTCCGCCGAAGCATGAAGAGCTGGACGAA 2159
QY 90 GlnAlaGlnPheLeuThrTrpAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
Db 2160 CTGATTCGATCTCCTCAGCAGCAATCCGATATTCGATCGAATGGTTCCGATCCGAC 2219
QY 110 GlnArgGlySerArgGluTyrAsnMetSerLeuGlyLysArgAlaValAlaValArg 129
Db 2220 AGGAAGGCCCGCGAGCTTCAACCTCGGACTATCTGACCGGAGGCCAATCCGTGTG 2279
QY 130 AsnTyrLeuLeuGlyLysGlyTleAsnGlnAlaSerValGluIleIleSerPheGlyLys 149
Db 2280 GATTACCTCAGCAGTGCTGCGATGAGCGCCGACAGGCTTACGTGGAAGGCTACGCTAAG 2339
QY 150 GlnArgPro 152
Db 2340 TCTGTCC 2348

RESULT 18
US-08-743-637B-11
; Sequence 11, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OUELLETTE & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 730 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Klebsiella pneumoniae
US-08-743-637B-11

Alignment Scores:
Pred. No.: 3.18e-09 Length: 730
Score: 136.50 Matches: 35
Percent Similarity: 47.41% Conservative: 20
Best Local Similarity: 30.17% Mismatches: 50
Query Match: 15.85% Indels: 11
DB: Gaps: 2

US-09-674-779B-2 (1-172) x US-08-743-637B-11 (1-730)

```
OY 66 LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAla 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 CTGAAGTCTGACGCTTCTGTTCACTTCAACAAAGCTACCTGAAACCGGAGTCAAG 427
OY 86 IleLeuAspGluGlnAlaGlnPheLeuThr-----AsnGlnThrAlaArgValLeu 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 GCTCTGGATCAGCGTCACTCACTCAGTCAGCAGCAATGATGATCCGAAACGCTTCC 487
OY 104 ValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyLys 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 GTTCTGGGCTACACCGCAGCGATCGGTTCCGAAAGCTTACACACGAGCGTCTGAGAA 547
OY 124 ArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGlu 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 CGTGTCACTGCGCTTGTGACTACCTGCTGCTAAAGGATCCCGCGTGGCAAAATCTCC 607
OY 144 IleIleSerPheGlyGluGlnArgProIleAlaPheGlyThrAsnGlu----- 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 GCTCGCGGCGATGGGTATCAACCGGTTACTGCGACACCTGTCGCAACGTAAGTAACT 667
OY 160 -----GluAlaTrpSerGlnAsnArgAlaGlnLeu 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 CGCGCTGCCCTGATGCTGCTGCTCGGATCGTGTGATGAGATC 715

RESULT 19
US-08-526-840B-11
Sequence 11, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
NUMBER OF SEQUENCES: 177
```

CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 730 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Klebsiella pneumoniae
US-08-526-840B-11

Alignment Scores:
Pred. No.: 3.18e-09 Length: 730
Score: 136.50 Matches: 35
Percent Similarity: 47.41% Conservative: 20
Best Local Similarity: 30.17% Mismatches: 50
Query Match: 15.85% Indels: 11
DB: Gaps: 2

US-09-674-779B-2 (1-172) x US-08-526-840B-11 (1-730)

```
OY 66 LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAla 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 CTGAAGTCTGACGCTTCTGTTCACTTCAACAAAGCTACCTGAAACCGGAGTCAAG 427
OY 86 IleLeuAspGluGlnAlaGlnPheLeuThr-----AsnGlnThrAlaArgValLeu 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 GCTCTGGATCAGCGTCACTCACTCAGTCAGCAGCAATGATGATCCGAAACGCTTCC 487
OY 104 ValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyLys 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 GTTCTGGGCTACACCGCAGCGATCGGTTCCGAAAGCTTACACACGAGCGTCTGAGAA 547
OY 124 ArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGlu 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 CGTGTCACTGCGCTTGTGACTACCTGCTGCTAAAGGATCCCGCGTGGCAAAATCTCC 607
OY 144 IleIleSerPheGlyGluGlnArgProIleAlaPheGlyThrAsnGlu----- 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 GCTCGCGGCGATGGGTATCAACCGGTTACTGCGACACCTGTCGCAACGTAAGTAACT 667
OY 160 -----GluAlaTrpSerGlnAsnArgAlaGlnLeu 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 CGCGCTGCCCTGATGCTGCTGCTCGGATCGTGTGATGAGATC 715

RESULT 20
US-08-836-500A-1
```

```

? Sequence 1 Application US/0836500A
? Patent No. 6197929
? GENERAL INFORMATION:
? APPLICANT: Binz, Hans
? APPLICANT: Baussant, Thierry
? APPLICANT: Haeuw, Jean-Francois
? APPLICANT: Nguyen Ngoc, Thien
? TITLE OF INVENTION: Carrier Protein Having an Adjuvant
? TITLE OF INVENTION: Effect, Immunogenic Complex Containing It, Process for
? TITLE OF INVENTION: Their Preparation, Nucleotide Sequence and Vaccines
? Patent No. 6197929
? NUMBER OF SEQUENCES: 8
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
? STREET: 180 N. Stetson, 2 Prudential Plaza, Suite
? STREET: 4700
? CITY: Chicago
? STATE: Illinois
? COUNTRY: U.S.A.
? ZIP: 60601
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentln Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/836,500A
? FILING DATE:
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: Katz, Martin L.
? REGISTRATION NUMBER: 25,011
? REFERENCE/DOCKET NUMBER: PIE1514P0180US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 312-616-5400
? TELEFAX: 312-616-5460
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1007 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1007
? US-08-836-500A-1
?
? Alignment Scores:
? Pred. No.: 5.17e-09 Length: 1007
? Score: 136.50 Matches: 35
? Percent Similarity: 47.41% Conservative: 20
? Best Local Similarity: 30.17% Mismatches: 50
? Query Match: 15.85% Indels: 11
? DB: 4 Gaps: 2
?
? US-09-674-779B-2 (1-172) x US-08-836-500A-1 (1-1007)
? QY 66 LeuprSerIeuValIyrPhesApPheasPserAspGluIleLysProGlnAlaIala 85
? ||| ||| ||| |||||:::||||:::
? Db 616 CTGAAGCTACGACTTCTGTTCACATTTCAACAAAGCTACCCTGAAMCGGAAGTGACGAG 675
? |||||:::||||:::
? QY 86 lIleuAspGluGlnAlaGlnPheLeurThr-----AsnGlnThrAlaArgValLeu 103
? |||||:::||||:::
? Db 676 GCCTGGATGAGCTGTACACTCAGCTAGCAGAACATGATGCCAAAAGACGGTCCGCTGT 735
? |||||:::||||:::
? QY 104 ValAlaGlyHisThrAspGluArgGlySerArgGluItyrAsmMetSerLeuGlyGluArg 123
? ||| |||||:::||||| ||||| ||| |||||
? Db 736 GTCTGGCGTACACCGCACCGCATCGGTCCGAAAGCTTACACACGACGACTGTCTGAGAA 795
? |||||:::||||| ||||| |||||
? QY 124 ArgAlaValAlaValaArgAsnTyrlLeuLeuGlyLySgIylleAsnGlnAlaSerValGlu 143
? ||||| ::||| ::|||::: |||||:::
? Db 796 CGRGCTAGTCCGTTGTGTGACTACCTGCTGTGCtTAAGGATCCCGGCTGCAAAATCTCC 855

```



```
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOHETICAL: yes
ORIGINAL SOURCE:
ORGANISM: Branchamella catarrhalis
STRAIN: 25240
IMMEDIATE SOURCE:
LIBRARY: genomic
CLONE: EMBL3 clone 5
CHROMOSOME/SEGMENT: pcd1
FEATURE:
LOCATION: CD gene region
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal sequence of encoded protein
LOCATION: -26 to -1
US-08-129-719-14

Alignment Scores:
Pred. No.: 1.36e-08 Length: 1727
Score: 136.00 Matches: 39
Percent Similarity: 49.07% Conservative: 14
Best Local Similarity: 36.11% Mismatches: 43
Query Match: 15.80% Indels: 12
DB: 1 Gaps: 5

US-09-674-779b-2 (1-172) x US-08-129-719-14 (1-1727)
QY 70 ValTyrPheaspheaspseraspgluileysprogin-----AlaIaIaIa 85
DB 1075 GTATTCCTTGTATGATTAATCAATCAACCAACCAACGTCGAAGAGTGTGAAG 1134
QY 86 IleleuaspgluInIaInpHeuThrThrasnInThraIaIaValIeValaIa 105
DB 1135 GTTCGTCGCAATGCGTGAATTC-----CCAAATGCAACTGCA-----ACCATGAA 1182
QY 106 GlyHisThrAspGlu-----ArgGlySerArgGluTyrAsnMetSerIeuGlyGlu 122
DB 1183 GGTCAACGATCAGCGGATTCAGACGCTCAAGTCAGCAACGCGCTCATCTGAA 1242
QY 123 ArgArgAlaValAlaValArgAsnTyrIleuLeuGlyLys---GlyIleAsnGlnAlaSer 141
DB 1243 GCTCGTCTTAATGCTTAATCAATCAATCAATGCTATCAACGATTGGTATCGCTCCAAACCGC 1302
QY 142 ValGluIleIleSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGluAla 161
DB 1303 CTAAATGCAAGTTGGTTATGCTTGTATGCTCTATCGTCCAAATACTACTGTAAGGT 1362
QY 162 TrpSerGlnAsnArgArgAlaGlu 169
DB 1363 AAAGCGATGAACCGCTGCTGTAAGAA 1386

RESULT 24
US-08-306-871-14
; Sequence 14, Application US/08306871
; Patent No. 5712118
; GENERAL INFORMATION:
; APPLICANT: Timothy F. Murphy
; TITLE OF INVENTION: Vaccine For Branchamella catarrhalis
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
; STREET: 1800 One Mt Plaza
; CITY: Buffalo
; STATE: New York
; COUNTRY: United States
; ZIP: 14203-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
; SOFTWARE: Wordperfect for Windows 5.1
; CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/306,871
FILING DATE: 20-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/129,719
FILING DATE: September 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Nelson, M. Bud
REGISTRATION NUMBER: 35,300
REFERENCE/DOCKET NUMBER: 11520.0053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 856-4000
TELEFAX: (716) 849-0349
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1727 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOHETICAL: yes
ORIGINAL SOURCE:
ORGANISM: Branchamella catarrhalis
STRAIN: 25240
IMMEDIATE SOURCE:
LIBRARY: genomic
CLONE: EMBL3 clone 5, SUBCLONE: pcd1
FEATURE:
LOCATION: CD gene region
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal sequence of encoded protein
LOCATION: -26 to -1
US-08-306-871-14

Alignment Scores:
Pred. No.: 1.36e-08 Length: 1727
Score: 136.00 Matches: 39
Percent Similarity: 49.07% Conservative: 14
Best Local Similarity: 36.11% Mismatches: 43
Query Match: 15.80% Indels: 12
DB: 1 Gaps: 5

US-09-674-779b-2 (1-172) x US-08-306-871-14 (1-1727)
QY 70 ValTyrPheaspheaspseraspgluileysprogin-----AlaIaIaIa 85
DB 1075 GTATTCCTTGTATGATTAATCAATCAACCAACCAACGTCGAAGAGTGTGAAG 1134
QY 86 IleleuaspgluInIaInpHeuThrThrasnInThraIaIaValIeValaIa 105
DB 1135 GTTCGTCGCAATGCGTGAATTC-----CCAAATGCAACTGCA-----ACCATGAA 1182
QY 106 GlyHisThrAspGlu-----ArgGlySerArgGluTyrAsnMetSerIeuGlyGlu 122
DB 1183 GGTCAACGATCAGCGGATTCAGACGCTCAAGTCAGCAACGCGCTCATCTGAA 1242
QY 123 ArgArgAlaValAlaValArgAsnTyrIleuLeuGlyLys---GlyIleAsnGlnAlaSer 141
DB 1243 GCTCGTCTTAATGCTTAATCAATCAATCAATGCTATCAACGATTGGTATCGCTCCAAACCGC 1302
QY 142 ValGluIleIleSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGluAla 161
DB 1303 CTAAATGCAAGTTGGTTATGCTTGTATGCTCTATCGTCCAAATACTACTGTAAGGT 1362
QY 162 TrpSerGlnAsnArgArgAlaGlu 169
DB 1363 AAAGCGATGAACCGCTGCTGTAAGAA 1386

RESULT 25
US-08-569-959-14
; Sequence 14, Application US/08569599
; Patent No. 5723862
; GENERAL INFORMATION:
; APPLICANT: Timothy F. Murphy
```


;; TITLE OF INVENTION: Vaccine For Brachnamella catarrhalis
;; NUMBER OF SEQUENCES: 52
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
;; STREET: 1800 One Mt Plaza
;; CITY: Buffalo
;; STATE: New York
;; COUNTRY: United States
;; ZIP: 14203-2391
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 Mb storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
;; SOFTWARE: Wordperfect for Windows 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/569,959
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: U.S. 08/129,719
;; FILING DATE: September 29, 1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Nelson, M. Bud
;; REGISTRATION NUMBER: 35,300
;; REFERENCE/DOCKET NUMBER: 11520.0053
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (716) 856-4000
;; TELEFAX: (716) 849-0349
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1727 nucleotides
;; TYPE: nucleic acid
;; STRANDEDNESS: single-stranded
;; TOPOLOGY: linear
;; MOLECULE TYPE: genomic DNA
;; HYPOTHEICAL: yes
;; ORIGINAL SOURCE:
;; ORGANISM: Brachnamella catarrhalis
;; STRAIN: 25240
;; IMMEDIATE SOURCE:
;; LIBRARY: genomic
;; CLONE: EMBL3 clone 5
;; FEATURE:
;; LOCATION: CD gene region
;; IDENTIFICATION METHOD: by experiment
;; NAME/KEY: signal sequence of encoded protein
;; LOCATION: -26 to -1
US-08-569-959-14
Alignment Scores:
Pred. No.: 1.36e-08 1727
Score: 136.00 Matches: 39
Percent Similarity: 49.078 Conservative: 14
Best Local Similarity: 36.118 Mismatches: 43
Query Match: 15.808 Indels: 12
Gaps: 5
DB: 1
US-09-674-779B-2 (1-172) x US-08-569-959-14 (1-1727)
QY 70 ValTYrPhAspPheAspSerAspGluLeuLysProGln-----AlaAlaAla 85
Db 1075 GTATCTCTTGAATTATGATCAATCAATCAACCAATACCGTGAAGAAGTGTCAAG 1134
QY 86 lLeLeAspGluGlnAlaGlnPheLeuThrAsnGlnPrrAlaArgValLeuAla 105
Db 1135 GTTGTGGCGCAATCGCGATTC-----CCAATGCAACTGCA-----ACCATGGAA 1182
QY 106 GlyHISThrAspGln-----ArgGlySerArgGluTyraSmetSerLeuGlyGlu 122
Db 1183 GGTCAAGCATCAGCGATTCAAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1242
QY 123 ArgArgAlaValAlaValArgAsnTyrlLeuLeuLys---GlylLeAsnGlnAlaSer 141
||||| |||||||:|:| |||

Db 1243 GCTCGCTAATGCTGTTAATCAATCTATCAACGAATTTGTATCCCTCCAAACCGC 1302
QY 142 ValGluLeuLeuSerPheGlyGluArgProIleAlaPheGlyThrAsnGluGlnAla 161
Db 1303 CTAATGCAAGTGTGTTATGCTTTGATGATGCTGCTCAATATCTACTGCTGAAGT 1362
QY 162 TrpSerGlnAsnArgArgAlaGlu 169
Db 1363 AAAGCATGAACCGCTGTAGAA 1386
RESULT 26
US-08-469-260A-80
; Sequence 80, Application US/08469260A
; Patent No. 6451578
; GENERAL INFORMATION:
; APPLICANT: JOHN N. SIMONS
; APPLICANT: TAMI J. PILOT-MATIAS
; APPLICANT: GEORGE J. DAMSON
; APPLICANT: GEORGE G. SCHLAUDER
; APPLICANT: SURESH M. DESAI
; APPLICANT: THOMAS P. LEARY
; APPLICANT: ANTHONY SCOTT MOERHOFF
; APPLICANT: JAMES C. ERKER
; APPLICANT: SHERI L. BUTIK
; APPLICANT: ISA K. MUSHAMMAR
; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
; NUMBER OF SEQUENCES: 716
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: 100 ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,260A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/424,550
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5527.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4268 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-260A-80
Alignment Scores:
Pred. No.: 0.192 4268
Score: 87.50 Matches: 38
Percent Similarity: 41.048 Conservative: 17
Best Local Similarity: 28.368 Mismatches: 48
Query Match: 10.168 Indels: 31
Gaps: 6
US-09-674-779B-2 (1-172) x US-08-469-260A-80 (1-4268)

QY	3	LeuHisIleGlnIleLeuAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGly	22
		:::: ::::	
Db	1686	CTCCGCTTTGGCGCTTGGAGTGGCTATGGCTATTATCCATGCACCTTTATTTGGCGCCACT	1745
QY	23	CysAlaAsnIysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThr	42
		::::	
Db	1746	TGTGTGGCGCGCTTCTGGTCTATTACATCAGTC-----CCATACGGGTGCTACT	1793
QY	43	GlyValIleIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValIysAlaLeu	62
		::::	
Db	1794	-----GTGCGCCCGAGTGGTGTACCAAGAGAAATCGTGAGAGAGTGT	1835
QY	63	AlaSerIysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleIysProGln	82
		::::	
Db	1836	GCATCATTCATATCCC-----TTGGAGGCCATG	1862
QY	83	AlaIleAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgVal	102
		::::	
Db	1663	GTTCGTCGCAATT---GACAAAGCTGAAGGTACATCAACACACA-----	1901
QY	103	LeuValAlaGlnIleHisThrAspGluThrArgGlySerArgGluTyrAsnMetSerLeuGlyGlu	122
		:::	
Db	1902	---ACTAGTCCTTTCCACATTGGAAACCCCTTGAAAAAACCTTACACACCTTTCTTGGGCGCT	1958
QY	123	ArgArgAla-----ValIleValAlaArgAsnTyrLeuLeuGly	134
		::::	
Db	1959	CATGCAGCTACATATCTTGCTACTATCAATAGATATTGCTGTGGC	2000

US-08-469-260A-11
Sequence 11, Application US/08469260A
Patent No. 6451578
GENERAL INFORMATION:
APPLICANT: JOHN N. SIMONS
APPLICANT: TAMT J. PILOT-MATIAS
APPLICANT: GEORGE J. DANSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MUEHRHOFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIRK
APPLICANT: ISA K. MUSHAWAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: POROMBSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5327.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SFO ID NO: 11


```

US-08-469-260A-393
Sequence 393, Application US/08469260A
Patent No. 6451578
GENERAL INFORMATION:
APPLICANT: JOHN N. SIMONS
APPLICANT: TAMI J. PILOT-MATTIAS
APPLICANT: GEORGE J. DAMSON
APPLICANT: GEORGE G. SCHLAUDER
APPLICANT: SURESH M. DESAI
APPLICANT: THOMAS P. LEARY
APPLICANT: ANTHONY SCOTT MOERHOFF
APPLICANT: JAMES C. ERKER
APPLICANT: SHERI L. BUIJK
APPLICANT: ISA K. MUSHAHMAR
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE
NUMBER OF SEQUENCES: 716
CORRESPONDENCE ADDRESS: D377/AP6D
ADDRESSEE: ABBOTT LABORATORIES
STREET: 100 ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,260A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/424,550
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FORDMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5527.PC.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 393:
SEQUENCE CHARACTERISTICS:
LENGTH: 9143 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..445
FEATURE:
NAME/KEY: CDS
LOCATION: 446..9037
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 9038..9143
US-08-469-260A-393
Alignment Scores:
Pred. No.: 0.606 Length: 9143
Score: 87.50 Matches: 38
Percent Similarity: 41.04% Conservative: 17
Best Local Similarity: 28.36% Mismatches: 48
Query Match: 10.16% Indels: 31
DB: 4 Gaps: 6
US-09-674-779B-2 (1-172) x US-08-469-260A-393 (1-9143)
3 Leuhtisileglllealllaalalaalaleuservaleuhtirphenetmthrly 22
|||||:::|||||||:::|||||:::|||||

```

```

Db      5141 CTCGCTGTTGGCGCTTGGAAGTGGCTAGAGCGTTATACGCAATTGACACTTTTGGCCCACT 5200
Qy      23  CysAlaAsnIlySerSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThr 42
Db      5201 TGTGTGGCGGGTGTGGTGTATATACATGATC-----CCATACGGGTCTACT 5248
Qy      43  GlyValIleIleYrThrGlyValAlaProLeuValAspAsnAspGluThrValIlyAsnAlaLeu 62
Db      5249 -----GTGCCCCAGTGGTGGACGACGAAGAAATCTGTAGGAGGTGT 5290
Qy      63  AlaSerIlyLeuProSerLeuValTyrPheAspPheAspSerAspGluIleIlyProGln 82
Db      5291 GCATCATCTATCTCC-----TTGAGGCCATG 5317
Qy      83  AlaAlaAlaIleLeuAspGlnGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgVal 102
Db      5318 GTTGTGTGCATC--GATTAAGCTGAAGAGTACATATACACCA----- 5356
Qy      103  LeuValAlaGlnIlyThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGlu 122
Db      5357 ---ACTAGTCTTTGACATTTGGAACCGCCCTTGAAAAACTTATACACTTTTGGGCGCT 5413
Qy      123  ArgArgAla-----ValAlaValAlaArgAsnTyrLeuLeuGly 134
Db      5414 CAGCAGCTACATCTCTGCTATCATATAGAGTATTGCTGTGCG 5455

RESULT 31
US-09-556-877-171
; Sequence 171, Application US/09556877
; Patent No. 6432816
; GENERAL INFORMATION:
; APPLICANT: Probst, Peter
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir
; APPLICANT: Filing, Steve
; APPLICANT: Malsomeuue, Jeff
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C5
; CURRENT APPLICATION NUMBER: US/09/556,877
; CURRENT FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 171
; LENGTH: 2895
; TYPE: DNA
; ORGANISM: Chlamydia
US-09-556-877-171

Alignment Scores:
Pred. No.: 0.81 Length: 2895
Score: 81.00 Matches: 39
Percent Similarity: 37.58% Conservative: 23
Best Local Similarity: 23.64% Mismatches: 75
Query Match: 9.41% Indels: 28
Gaps: 7

US-09-674-779B-2 (1-172) x US-09-556-877-171 (1-2895)
Qy      8  AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnIlySer 27
Db      829  GCTTCTGATGAGGAGGACATTAAGTAACTACTCCCTACAGATGTTACAGGCAATCGT--- 885
Qy      28  ThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleYrThr 47
Db      886 ---GGTAGAGATCTTTTATAGTACAAATATACAAAAAATTTATGGCGGAGCTATTACGCT 942
Qy      48  GlyValAlaProLeuValAspAsnAspGluThr-----ValIlyAlaLeuAlaSerIly 65
Db      943  CCGTAGTACCTTGGATGATATAGCCCTACTCTTATTATACAAATATGCCCAATATAT 1002
Qy      66  LeuProSerLeuValTyrPheAspPheAspSerAspGluIleIlyProGlnAlaAlaAla 85

```


GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1
Alignment Scores:
Pred. No.: 359 Length: 80161
Score: 77.50 Matches: 32
Percent Similarity: 33.97% Conservative: 21
Best Local Similarity: 20.51% Mismatches: 48
Query Match: 9.00% Indels: 55
Gaps: 5
US-09-674-779B-2 (1-172) x US-09-036-987A-1 (1-80161)
QY 29 SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGly 48
Db 5864 ACCGGAGGTGATGTCGACCTCTTGGC-----GCAGGTGTGTCACAGCGCGG 5911
QY 49 ValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSer 68
Db 5912 CTCGACACGACGATCCGCTCCACGATCGACGACGA----- 5947
QY 69 LeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
Db 5948 -----CTCGAT 5953
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 5954 GATGGAGCCAAATTCACAGGCTGCGGTCACACAGCAGCCACCTCGCGCTGGATGTG 6013
QY 109 AspGluArg----- 111
Db 6014 ATGGCAGGATAGAGAGCTGACCGTTGTCAGGTCCGCGAAGGTTCGGGTACAGCCGTC 6073

QY 112 -----GlySerAlaGluTyrAsnMetSerLeuGlyValArgAlaVal----- 126
Db 6074 GTAGGTGGATCCACCGCATCACAGTTGTCCTTCGACGAGGACAGCTCGTCTTTCG 6133
QY 127 AlaValArgAsnTyrLeuLeuGlyLysGlyTyrLeuGlnAlaSerValGluIleLeuSer 146
Db 6134 GATCTCCGCAATTCGTCAGCGGCGAAGGGGTTCGCCAGGT-----GTGCGCCAG 6184
QY 147 PheGlyGluGlnArgProIleAlaPheGlyThrAsnGluGlnAlaTrp 162
Db 6185 CATGATGCTGCGCTCGCTCCGTCGACGCGGCGCTTATCGGTCGTG 6232
RESULT 37
US-09-370-700-1
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
EARLIER FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 80161
TYPE: DNA
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1
Alignment Scores:
Pred. No.: 359 Length: 80161
Score: 77.50 Matches: 32
Percent Similarity: 33.97% Conservative: 21
Best Local Similarity: 20.51% Mismatches: 48
Query Match: 9.00% Indels: 55
Gaps: 5
US-09-674-779B-2 (1-172) x US-09-370-700-1 (1-80161)
QY 29 SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGly 48
Db 5864 ACCGAGGTGATGTCGACCTCTTGGC-----GCAGGTGTGTCACAGCGCGG 5911
QY 49 ValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSer 68
Db 5912 CTCGACACGACGATCCGCTCCACGATCGACGACGA----- 5947
QY 69 LeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
Db 5948 -----CTCGAT 5953
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 5954 GATGGAGCCAAATTCACAGGCTGCGGTCACACAGCAGCCACCTCGCGCTGGATGTG 6013
QY 109 AspGluArg----- 111
Db 6014 ATGGCAGGATAGAGAGCTGACCGTTGTCAGGTCCGCGAAGGTTCGGGTACAGCCGTC 6073
QY 112 -----GlySerAlaGluTyrAsnMetSerLeuGlyValArgAlaVal----- 126
Db 6074 GTAGGTGGATCCACCGCATCACAGTTGTCCTTCGACGAGGACAGCTCGTCTTTCG 6133

```
QY    127 AValAargsnrlyrrleuaglylsyglVlIasnmInlaSerValguilleSer 146
      :::::::::::::::::::: ||:::||::| |::|
Db     6134 GAATCCCGCATTTTCGACAGCGGAAGGGGTGCCAGGT-----GTGCCTCAAG 6184

QY    147 PheGIglnGluaArgProIleaIApheGIYThraSngluAlatrp 162
      ||||| |::| |::|
Db     6185 CATGATGCTCAGTCCGTCCTCCGAGACGGCGCTTGATGCGGTCTGG 6232

RESULT 38
US-09-103 -840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294528
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103, 840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:          1,59e+05           Length:       4403765
Score:             77.00              Matches:        39
Percent Similarity: 41.96%            Conservative:   21
Best Local Similarity: 27.27%         Mismatches:    58
Query Match:       8.94%               Indels:        25
                                      Gaps:          7
DB:                                4                                     7

US-09-674-779B-2 (1-172) x US-09-103-840A-2 (1-4403765)
QY      8 AAlaAlaAlaAlaAlaLeuSeRValLeuthRhPheMeThrgly----- 22
      ||||||| |::| :::::: |||||
Db     3915272 GCCCGCCGACAAGCGTGCGATCACCGCCTGTCCGGCGCGCGCGCTCGGCCG 3915213

QY     23 -----CYsaIsAsnlSySerThrserGlnValMetValAlaProSaMnLaProthr 39
      ||||||| |::| :::::: |||||
Db     3915212 TAGCGCGCTTCCGCCCAATT-----GTGATTGTTCGGCGGCGCGGCACC 3915171

QY      40 GLYTyrThrGlyValLIetYTrHrGlyValAlaProLeuValaSpaaSnaspGLutHrVal 59
      ||| :::::: ||| ||||| |::| :::
Db     3915170 GCAGTAGACGCCGCAATGTCTTC---GGCGCCGCACCCAGTCGTAAGAAGCGGCACAGATCGAC 3915114

QY      60 LysAlaLeuAlaSerLysLeuProSerLeuValTyrrPheaSprPheaSprerAspGUlle 79
      ||||| |::| :::::: ||||| |::|
Db     3915113 CGCGTGTCCGCGCAcAGATGTGTAAcCTGTGTCCAATT-----CTG 3915072

QY      80 LysPROGlnAIalAAlaAlIIelLeuASPgluglnAlaGlnPheLeuThrhThrasnGlnIthr 99
      ||||||| |::| :::::: |||
Db     3915071 GCGTCCCCGCGTCCGCCGGAAGTCACAGCTCAGG--TTCATCGTCTACGTCGCGAG 3915015

QY     100 AlaArGvalLeuValAlaGlyHIsthrAspGuArgLySerARygGluTYrrAsmMelSer 119
      ::::::: ||| ::: ||| ||| ::: |||
Db     3915014 GTGACGCGGTGTCAccCGCCCaCATGaTGAAGGCGGCTTCAcAGCGCGAc--GGCAGCTGC 3914958

QY     120 LeuGIglnUARArgVALa--ValAla---VaIRAsntYrrleuGnGLYtSYllEaS 138
      ||||| |::| |::| ||| :::||| |::|
Db     3914957 CTGGGATCCCAcCAacGACTCACCGcGAGCcTGCggGAgTACTTTGCTGTGGATCCGGA 3914898

QY     138 nGlnAla 140
```

```
DB      3914897 ACAGAGC 3914891
RESULT 39
US-09-453-702B-79
; Sequence 79, Application US/09453702B
; Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
          Burland, Valerie
          Perna, Nicole T.
          Plunkett, Guy
          Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Charles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
SEQUENCE CHARACTERISTICS:
LENGTH: 38155
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-453-702B-79
Alignment Scores:
Pred. No.:           349             Length:       38155
Score:              74.00            Matches:        40
Percent Similarity: 39.89%           Conservative:   31
Best Local Similarity: 22.47%         Mismatches:    57
Query Match:        8.59%            Indels:         50
DB:                  4                Gaps:           8
US-09-674-779B-2 (1-172) x US-09-453-702B-79 (1-38155)
QY      3 LeunHstlEgInIIealAalaAlaAlaAlaAlaAlaAlaLeSeRValLeuThrPhemeThrgly 22
      |||::||| | | | | | | | | | | | | | | | | :| | | |
Dbl     35145 CTGAACAAGAATATTGTCGCCACAGAAAGCGCCGCAGAAATCGCTGAACAGTACACGGCG 35204
QY      23 CysAlasAnLysSerThrSergIlnValMetValAlaProAsnAlaProlTrglYTyrrH 42
      ::::::| | | | | | | | | | | | | | | | :| | | |
Dbl     35205 GTGGGCTCACGTCTGATGAGCGGCGCTG-----GGGCTGTTGGTGCGTACC 35255
QY      43 GIyVALtIEryTrhgLyVAlAla-----ProleuVAlaspnsApGIu----- 57
      |||:::||| | | | | | | | | | | | | | | | | | | |
Dbl     35236 GGACTGGTGATGCTGGGGGCTGCAGCATGTGTACACGCTGTACCAGAAATCAGAGCAAGGCC 35315
```



```

QY 58 -----ThrVallylsAlaLeuAlaSerlyLeuProser 68
Db 35316 AGGAGTCTGCGCCAGTATGACATGACATGAATGAAATGCGCATTAACCGCTCA 35375
QY LeuValTyrPheaspPheaspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
Db 35376 ATGCTCTTGGCTGAACCTCAGATATATGAAGCAGACGCGCGCGC-----CTGACA 35429
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 35430 GAGCAGAACCGCGTGATT----- 35447
QY 109 AspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaVal 128
Db 35448 GATGAACAGCGCCAGTGGGTGAATCCCTG-----CAGGAAAAAGCGCATCCATT 35498
QY 129 ArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db 35499 CAGGATGTGCTTCCCGGCGT----- 35519
QY 149 GluGluArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArg 166
Db 35520 GAAGACCGCTGCTGGCGCTTA---ATTGCTCAGCAGCGCGGAGAGCATTAAG 35570

RESULT 40
US-09-453-702B-137/C
; Sequence 137, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-Dec-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296, 95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48908
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 137:
US-09-453-702B-137
Alignment Scores:

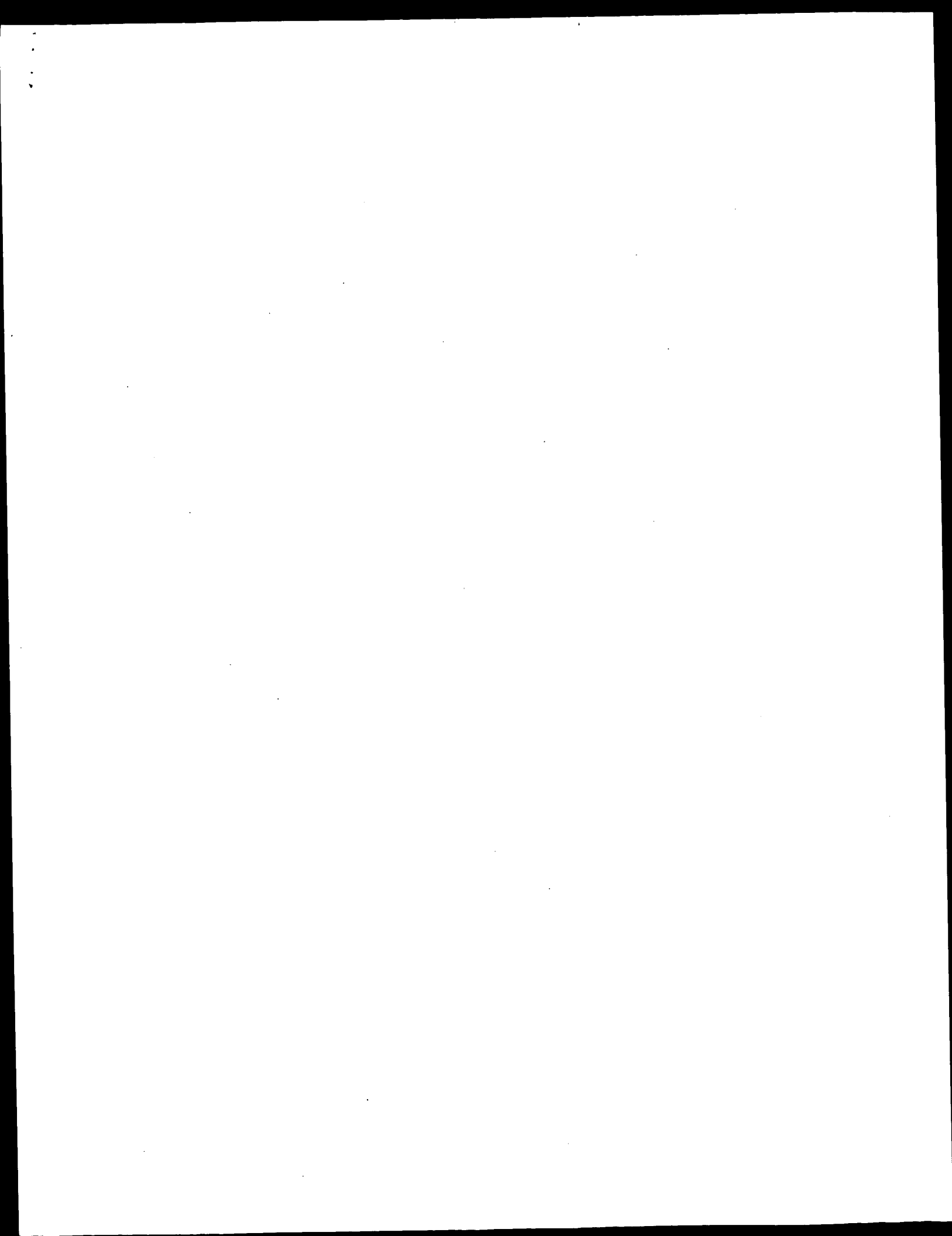
```

```

Pred. No.: 507
Score: 74.00
Percent Similarity: 39.89%
Best Local Similarity: 22.47%
Query Match: 8.59%
DB: 4
Gaps: 8
US-09-674-779b-2 (1-172) x US-09-453-702B-137 (1-48908)
QY 3 LeuHisIleGlnIleAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGly 22
Db 11825 CTGAACAGAAATATTTGCTGCCAGAGAGCGCCGAGATGCGTGAACAGTACACGGCG 11766
QY 23 CysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThr 42
Db 11765 GTGGGCTCAGCTGTCATGATGCGGCGT-----GGCTGCTGTGGTGGCTGACCC 11715
QY 43 GlyValIleIleTyrThrGlyValAla-----ProLeuValAspAsnAspGlu----- 57
Db 11774 GGACTGGTGAATGCTGGGGGCTGCGACATGTGACACGCTGTACAGAAATCAGACGAGCGCC 11655
QY 58 -----ThrVallylsAlaLeuAlaSerlyLeuProser 68
Db 11654 AGGAGTCTGCGCCAGTATGACATGACATGAATGCGCATTAACCGCTCA 11595
QY 69 LeuValTyrPheaspPheaspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
Db 11594 ATGCTCTTGGCTGAACCTCAGATATATGAAGCAGAACGCGCGCGC-----CTGACA 11541
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
Db 11540 GAGCAGAACCGCGTGATT----- 11523
QY 109 AspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaVal 128
Db 11522 GATGAACAGCGCCAGTGGGTGAATCCCTG-----CAGGAAAAAGCGCATCCATT 11472
QY 129 ArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db 11471 CAGGATGTGCTTCCCGGCGT----- 11451
QY 149 GluGluArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArg 166
Db 11450 GAAGACCGCTGCTGGCGCTTA---ATTGCTCAGCAGCGCGGAGAGCATTAAG 11400

Search completed: July 6, 2003, 13:40:31
Job time : 1768 secs

```



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:52:45 ; Search time 139 Seconds

(without alignments)
1931.586 Million cell updates/sec

Title: US-09-674-779b-2

Sequence: 1 MMHIIQIAAAALSVLTFM.....IAFGNNEASQNRRLASY 172

Scoring table: BIOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Command line parameters:

-MODE=frame.p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09674779/unat.06072003.121615.9891/app_query.fasta.1.327
-DB=Published_Applications_NA -OFM=fastap -SUFFIX=p2n.rmpb -MIMATCH=0.1
-LOOPEXT=0 -LOOPEXT-0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62
-TRANS=human40.cdi -LIST=100 -DOCALL=200 -THR.SCORE=pcr -THR.MAX=100
-THR.MIN=0 -ALIGN=40 -MODE=LOCAL -OUT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO_MAP -LARGEQUERY -NEG.SCORES=0 -NAIT -DSPLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOR=6 -DELEXT=7

Database: Published_Applications_NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	264	30.7	1830121	9 US-10-329-980-1
2	204.5	23.8	779	10 US-09-747-348-1
3	176.5	20.5	1830121	9 US-10-329-980-1
4	160	18.6	984	9 US-10-216-338A-5

5	149	17.3	678	9 US-10-156-761-3232	Sequence 3332, Ap
6	149	16.8	861	9 US-10-216-338A-7	Sequence 1, Appl
7	145	15.9	730	10 US-09-452-559-11	Sequence 7, Appl
8	136.5	15.9	1035	10 US-10-169-953-1	Sequence 11, Appl
9	136.5	15.7	640681	10 US-09-790-988-1	Sequence 1, Appl
10	135.5	15.7	561	10 US-09-974-300-1902	Sequence 1902, Ap
11	135.5	13.2	1143	9 US-09-998-278-25	Sequence 25, Appl
12	107	12.4	669	10 US-09-974-300-1889	Sequence 1889, Ap
13	101.5	11.8	717	9 US-10-010-160-11	Sequence 11, Appl
14	93.5	10.9	4268	8 US-08-424-550B-80	Sequence 80, Appl
15	87.5	10.2	8912	8 US-08-424-550B-11	Sequence 11, Appl
16	87.5	10.2	9143	8 US-08-424-550B-390	Sequence 390, Ap
17	87.5	10.2	9143	8 US-08-424-550B-393	Sequence 393, Ap
18	87.5	10.2	9399	10 US-09-742-659-1	Sequence 1, Appl
19	87.5	10.2	8033	10 US-09-070-927A-121	Sequence 121, Ap
20	87	10.1	2816	10 US-09-886-468-8	Sequence 8, Appl
21	82	9.5	4224	10 US-09-841-132-486	Sequence 171, Ap
22	82	9.4	2895	10 US-09-841-132-171	Sequence 183, Ap
23	81	9.4	2934	10 US-09-841-132-183	Sequence 183, Ap
24	81	9.3	1086	12 US-10-007-693-58	Sequence 486, Ap
25	80	9.3	1260	9 US-10-156-761-1993	Sequence 171, Ap
26	80	9.3	9025608	9 US-10-156-761-1	Sequence 1993, Ap
27	80	9.1	315	10 US-09-974-300-6221	Sequence 1, Appl
28	78.5	9.0	1634	12 US-10-007-693-43	Sequence 43, Appl
29	77.5	9.0	1762	12 US-10-062-254-323	Sequence 323, Ap
30	77.5	9.0	2631	10 US-10-156-761-1454	Sequence 6208, Ap
31	77.5	9.0	1317	10 US-09-815-242-9978	Sequence 9978, Ap
32	76.5	8.9	1317	10 US-09-815-242-9978	Sequence 60, Appl
33	76	8.8	2868	10 US-09-960-253-145	Sequence 145, Ap
34	76	8.8	10300	10 US-10-156-761-15966	Sequence 145, Ap
35	75.5	8.8	465	9 US-09-815-242-4642	Sequence 5966, Ap
36	75.5	8.8	2382	10 US-09-815-242-4642	Sequence 4642, Ap
37	75.5	8.8	2409	10 US-09-815-242-4642	Sequence 4642, Ap
38	75	8.7	1335	9 US-10-156-761-1454	Sequence 1455, Ap
39	75	8.7	1707	10 US-09-815-242-4085	Sequence 4085, Ap
40	75	8.7	11575	10 US-09-938-956-5	Sequence 2960, Ap
41	74.5	8.7	3990	9 US-10-156-761-2960	Sequence 5966, Ap
42	74	8.6	605	9 US-09-796-692-6743	Sequence 6743, Ap
43	74	8.6	605	9 US-10-040-662-6743	Sequence 6743, Ap
44	74	8.6	1387	9 US-10-198-846-13428	Sequence 13428, A
45	74	8.6	1860	9 US-10-156-761-2778	Sequence 2778, Ap
46	74	8.6	2265	9 US-10-037-770-582	Sequence 582, Ap
47	74	8.6	3180	12 US-10-044-090-663	Sequence 663, Ap
48	74	8.6	3534	9 US-10-128-714-2453	Sequence 2453, Ap
49	74	8.6	3540	9 US-10-128-714-7453	Sequence 7453, Ap
50	74	8.6	3877	9 US-10-128-714-1493	Sequence 1493, Ap
51	74	8.6	3877	9 US-10-128-714-6493	Sequence 6493, Ap
52	74	8.6	5877	9 US-10-128-714-493	Sequence 493, Ap
53	74	8.6	5877	9 US-10-128-714-5493	Sequence 5493, Ap
54	74	8.6	38155	9 US-10-114-170-79	Sequence 79, Appl
55	74	8.6	48908	9 US-10-114-170-137	Sequence 137, Appl
56	73.5	8.5	4494	9 US-09-918-995-24922	Sequence 24922, A
57	73.5	8.5	1707	9 US-09-738-626-2679	Sequence 2679, Ap
58	73.5	8.5	1720	9 US-10-128-714-1561	Sequence 1561, Ap
59	73.5	8.5	1720	9 US-10-128-714-6581	Sequence 6581, Ap
60	73.5	8.5	2787	10 US-09-815-242-6320	Sequence 6320, Ap
61	73.5	8.5	3720	9 US-10-128-714-581	Sequence 581, Ap
62	73.5	8.5	3720	9 US-10-128-714-5581	Sequence 5581, Ap
63	73.5	8.5	6791	9 US-09-771-035A-1	Sequence 31, Appl
64	73.5	8.5	13856	7 US-08-781-986A-31	Sequence 1, Appl
65	73.5	8.5	3309400	9 US-09-738-626-1	Sequence 330, Ap
66	73	8.5	4205	9 US-10-098-841-330	Sequence 480, Ap
67	73	8.5	4226	9 US-10-037-270-480	Sequence 4513, Ap
68	73	8.5	4473	10 US-09-815-242-4513	Sequence 8064, Ap
69	73	8.5	4509	10 US-09-815-242-8064	Sequence 944, Ap
70	73	8.5	4560	10 US-09-738-626-944	Sequence 80, Appl
71	73	8.5	5659	12 US-10-002-600-79	Sequence 79, Appl
72	73	8.5	5763	12 US-10-002-600-79	Sequence 4, Appl
73	73	8.5	6002	9 US-10-071-179-4	Sequence 252, Appl
74	73	8.5	12848	9 US-10-114-170-252	Sequence 56, Appl
75	73	8.5	30246	7 US-08-781-986A-56	Sequence 1, Appl
76	73	8.5	3309400	9 US-09-738-626-1	Sequence 4240, Ap
77	72.5	8.4	915	9 US-10-156-761-4240	

```
c 78 72.5 8.4 1674 9 US-10-156-761-3205
79 72.5 8.4 2553 9 US-10-156-761-2383
80 72.5 8.4 10610 9 US-09-994-412-3
81 72.5 8.4 11517 10 US-09-901-106-1
82 72.5 8.4 885 9 US-10-156-761-1076
83 71.5 8.3 770 9 US-10-198-846-12573
84 71.5 8.3 1323 9 US-10-156-761-6130
85 71.5 8.3 1521 9 US-10-156-761-6926
86 71.5 8.3 7516 10 US-09-917-800A-1664
87 71.5 8.2 514 9 US-09-918-995-25705
88 71.5 8.2 546 9 US-10-114-666-9
89 71.5 8.2 546 10 US-09-895-828-9
90 71.5 8.2 797 10 US-09-880-107-3806
91 71.5 8.2 886 9 US-09-996-634-119
92 71.5 8.2 886 9 US-09-997-181-119
93 71.5 8.2 886 9 US-09-997-182-119
94 71.5 8.2 1389 9 US-10-156-761-968
95 71.5 8.2 1391 10 US-09-925-301-66
96 71.5 8.2 4056 9 US-10-156-761-2317
97 71.5 8.2 6694 9 US-10-074-095-700
98 71.5 8.2 6694 10 US-09-764-860-700
99 71.5 8.2 7684 10 US-09-070-927A-16
100 71.5 8.2 8688 9 US-09-938-842A-1748
```

ALIGNMENTS

```
Sequence 3205, Ap
Sequence 2383, Ap
Sequence 3, Appl
Sequence 1, Appl
Sequence 1076, Ap
Sequence 12573, A
Sequence 6130, Ap
Sequence 6926, Ap
Sequence 1664, Ap
Sequence 25705, A
Sequence 9, Appl
Sequence 8, Appl
Sequence 3806, Ap
Sequence 119, App
Sequence 119, App
Sequence 119, App
Sequence 968, App
Sequence 66, Appl
Sequence 2317, Ap
Sequence 700, App
Sequence 700, App
Sequence 16, Appl
Sequence 1748, Ap
```

```
RESULT 1
US-10-329-960-1/c
Sequence 1, Application US/10329960
Publication No. US20030099277A1
```

```
GENERAL INFORMATION:
```

```
APPLICANT: Fleischmann et al.
```

```
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
```

```
TITLE OF INVENTION: Thereof, and Uses Thereof
```

```
FILE REFERENCE: P186P1
```

```
CURRENT APPLICATION NUMBER: US/10/329,960
```

```
PRIOR FILING DATE: 2003-01-02
```

```
PRIOR APPLICATION NUMBER: US 09/643,990
```

```
PRIOR FILING DATE: 2000-08-23
```

```
PRIOR APPLICATION NUMBER: US 08/487,429
```

```
PRIOR FILING DATE: 1995-06-07
```

```
PRIOR APPLICATION NUMBER: US 08/426,787
```

```
PRIOR FILING DATE: 1995-04-21
```

```
NUMBER OF SEQ ID NOS: 1
```

```
SOFTWARE: PatentIn version 3.1
```

```
SEQ ID NO 1
```

```
LENGTH: 1830121
```

```
TYPE: DNA
```

```
ORGANISM: Haemophilus Influenzae
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (4747)..(4747)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (9921)..(9921)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (10150)..(10150)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
```

```
NAME/KEY: misc_feature
```

```
LOCATION: (36543)..(36543)
```

```
OTHER INFORMATION: n equals a, t, g or c
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (53569)..(53569)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (13910)..(13910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
```

```
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

Alignment Scores:

Pred. No.:	6.02e-23	Length:	1830121
Score:	264.00	Matches:	51
Percent Similarity:	68.93%	Conservative:	20
Best Local Similarity:	49.51%	Mismatches:	32
Query Match:	30.66%	Indels:	0
DB:	9	Gaps:	0

US-09-674-779b-2 (1-172) x US-10-329-960-1 (1-1830121)

```
Qy 70 ValtyrPheaspPheaspSeraspGluIlelyspProglAlaAlaIleleuaspGlu 89
Db 401840 GTTATTTCGTTTGAATAATATGACATTACTGTGATACGTTCAATCTTAGACGG 401781
Qy 90 GluAlaGlnPheleuthrThrAsnGlnThrAlaArgValleuValAlaGlyIsthrasp 109
Db 401780 CAGCTGCATATTTAAATGCAACGCCAGCTGTAAAGTATTAAGTAAAGTAAAGTAA 401721
Qy 110 GluArgGlySerArgGlyuTyraAsnMetSerLeuGlyGluArgGAlaValAlaArg 129
Db 401720 GAACGTGTACACAGATACATACATCGCATTTAGGCCACGTGTGCATGCACTTAA 401661
Qy 130 AsnTyrlleuLeuGlylysglyIleasnGlnAlaSerValGluIleIleSerPheGlyGlu 149
Db 401660 GGTATTATTAGCTGTAAAGGTGTGTGATGCTGTAAATTTAGGCACACTATCTTACG 401601
Qy 150 GluArgProIleAlaPheGlyThrAsnGlnGluAlaIleIlePheGlnAsnArgAlaGlu 169
Db 401600 GAAAAACCTGCAGTATTAAGTCAATGATGAAGCTGATATTTAAAAACCGTGTGAC 401541
Qy 170 LeuSerTyrr 172
Db 401540 TTAGCGTAC 401532
```

RESULT 2

```
US-09-747-348-1
; Sequence 1, Application US/09747348
; Patent No. US20020123067A1
; GENERAL INFORMATION:
; APPLICANT: Andrew D. MURDIN, Raymond P. COWEN, Joe WANG, Pamela DUNN
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses th
; FILE REFERENCE: 77813-37
; CURRENT APPLICATION NUMBER: US/09/747,348
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/171,525
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 779
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(676)
US-09-747-348-1
```

Alignment Scores:

Pred. No.:	1.14e-19	Length:	779
Score:	204.50	Matches:	46
Percent Similarity:	58.82%	Conservative:	24
Best Local Similarity:	38.66%	Mismatches:	46

```

1 NAME/KEY: misc_feature
2 LOCATION: (.36551)..(.36551)
3 OTHER INFORMATION: n equals a, t, g or c
4 FEATURE:
5 NAME/KEY: misc_feature
6 LOCATION: (.36636)..(.36636)
7 OTHER INFORMATION: n equals a, t, g or c
8 FEATURE:
9 NAME/KEY: misc_feature
10 LOCATION: (.40808)..(.40810)
11 OTHER INFORMATION: n equals a, t, g or c
12 FEATURE:
13 NAME/KEY: misc_feature
14 LOCATION: (.44416)..(.44416)
15 OTHER INFORMATION: n equals a, t, g or c
16 FEATURE:
17 NAME/KEY: misc_feature
18 LOCATION: (.44975)..(.44975)
19 OTHER INFORMATION: n equals a, t, g or c
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: (.44905)..(.44905)
23 OTHER INFORMATION: n equals a, t, g or c
24 FEATURE:
25 NAME/KEY: misc_feature
26 LOCATION: (.44975)..(.44975)
27 OTHER INFORMATION: n equals a, t, g or c
28 FEATURE:
29 NAME/KEY: misc_feature
30 LOCATION: (.45732)..(.45732)
31 OTHER INFORMATION: n equals a, t, g or c
32 FEATURE:
33 NAME/KEY: misc_feature
34 LOCATION: (.45732)..(.45732)
35 OTHER INFORMATION: n equals a, t, g or c
36 FEATURE:
37 NAME/KEY: misc_feature
38 LOCATION: (.47036)..(.47036)
39 OTHER INFORMATION: n equals a, t, g or c
40 FEATURE:
41 NAME/KEY: misc_feature
42 LOCATION: (.51334)..(.51334)
43 OTHER INFORMATION: n equals a, t, g or c
44 FEATURE:
45 NAME/KEY: misc_feature
46 LOCATION: (.51602)..(.51602)
47 OTHER INFORMATION: n equals a, t, g or c
48 FEATURE:
49 NAME/KEY: misc_feature
50 LOCATION: (.51786)..(.51786)
51 OTHER INFORMATION: n equals a, t, g or c
52 FEATURE:
53 NAME/KEY: misc_feature
54 LOCATION: (.55369)..(.55369)
55 OTHER INFORMATION: n equals a, t, g or c
56 FEATURE:
57 NAME/KEY: misc_feature
58 LOCATION: (.65309)..(.65309)
59 OTHER INFORMATION: n equals a, t, g or c
60 FEATURE:
61 NAME/KEY: misc_feature
62 LOCATION: (.65313)..(.65313)
63 OTHER INFORMATION: n equals a, t, g or c
64 FEATURE:
65 NAME/KEY: misc_feature
66 LOCATION: (.80024)..(.80024)
67 OTHER INFORMATION: n equals a, t, g or c
68 FEATURE:
69 NAME/KEY: misc_feature
70 LOCATION: (.100091)..(.100091)
71 OTHER INFORMATION: n equals a, t, g or c
72 FEATURE:
73 NAME/KEY: misc_feature

```

```
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
```

```
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 2,63e-10
Score: 176.50
Percent Similarity: 47.77%
Best Local Similarity: 30.57%
Query Match: 20.50%
DB: 9
Gaps: 4

US-09-674-779b-2 (1-172) x US-10-329-960-1 (1-1830121)

QY 35 ProasAlaProthGlyTyr-----ThrGlyValIleTyr 46
DB 1235275 CCAAAATACCGCAGCTTAACCTACCAACCTTGATGCTGTATCAATGCGGATATTCCTTAC 1235334
QY 47 -----ThrGlyValAlaProLeuValAlaSpasnaSpGluThrValLysAlaLeuAla 63
DB 1235335 CGTTTGGTCAAGGCGGACCAAGTGTTCACACACCTGAAGTTGTAGCAAACTTTC 1235394
QY 64 SerLysIeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAla 83
DB 1235395 AGC--TTAACTGTGATGATCACTTGCATTTGTAAGCAAACTTAAACCTCAAGCA 1235451
QY 84 AlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeu 103
DB 1235452 CAAGCAACATTAAGACAGCATCTATGCGGAATGTCAACAAGTTAAAGTCAAAAGTACT 1235511
QY 104 ValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArg 123
DB 1235512 GTTGTGTGTTCACCTGACCCGATTTGTCAGCCGTTCAACGTAANAACCTTCTCAAGAA 1235571
QY 124 ArgAlaValAlaValAlaGAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGlu 143
DB 1235572 CGTCAAGATTCAGTAGCTACTTGTGCTAAAGTGTTCAGACGACGCAATCTCT 1235631
QY 144 IleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu----- 160
DB 1235632 GCAACTGGCTAAGGTAAGCAAAACCGTAAGTGTGCAACTGTGTGCAAGTTAAGGCT 1235691
QY 161 -----AlaTyrSerGlnAsnArgArgAlaGluLeuSer 171
DB 1235692 CGTAAAGCGCTTATCGCTGTGTTGCTCCAGACCGTCGCTGAAGATCGCG 1235742

RESULT 4
US-10-216-338A-5
: Sequence 5, Application US/10216338A
: Publication No. US20030113757A1
: GENERAL INFORMATION:
: APPLICANT: CzaJka, John
: TITLE OF INVENTION: Rapid and Specific Detection of Campylobacter
: FILE REFERENCE: MD10830SNA
: CURRENT APPLICATION NUMBER: US/10/216,338A
: PRIOR FILING DATE: 2003-01-14
: PRIOR APPLICATION NUMBER: 60/310,882
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 5
: LENGTH: 984
: TYPE: DNA
: ORGANISM: Campylobacter coli
```

US-10-216-338A-5

Alignment Scores:

Pred. No.:	4.44e-13	Length:	984
Score:	160.00	Matches:	33
Percent Similarity:	52.94%	Conservative:	21
Best Local Similarity:	32.35%	Mismatches:	48
Query Match:	18.58%	Indels:	0
DB:	9	Gaps:	0

US-09-674-779B-2 (1-172) x US-10-216-338A-5 (1-984)

```

QY 71 TyrPheasphepserspqluileysproglalaalaileuaspqlu 90
    ::::: ||||| ::::: ||| ::::: |||
DB 664 CATTTGGTTTGTATAGTATCAATCCAGCTTTGAGAAAAATCAAGAAAT 723
    ||||| ||| ||| ::::: ||||| |||||
QY 91 AlaGlnPheLeuThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGlu 110
    ||||| ||| ||| ::::: ||||| |||||
DB 724 GCTCACTTTATGATGAAATGCAAGATATGATCTATTATTAGAGGCTACTGCTATAT 783
    ||||| ||| ||| ::::: ||||| |||||
QY 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsn 130
    ||||| ||||| ||||| ||||| ||||| |||||
DB 784 ATAGGCTCCAGACCAATCAAACTTCAGAAAGAGGCTGAAAGCCGTGCAGAA 843
    ||| ||||| ||||| ||||| ||||| |||||
QY 131 TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleSerPheGlyGlu 150
    ||| ||||| ||||| ||||| ||||| |||||
DB 844 GAACCTGAAAAATTTGGTGTAGTAAAGATCGTATCCAGACGTTGGTATGTCAGAT 903
    ||| ||||| ||||| ||||| ||||| |||||
QY 151 ArgProIleAlaPheGlyThrAsnGluAlaIleTrpSerGlnAsnArgAlaGluLeu 170
    ::::: ||||| ::::: ||||| ||||| |||||
DB 904 AAACCTCGCTCAGAAATGAGACCAAGAGGTAGAGAGATTAACAGAGTGTGCT 963
    ::::: ||||| ::::: ||||| ||||| |||||
QY 171 SerTyr 172
    ::::: |||||
DB 964 AAATTT 969
    ::::: |||||

```

RESULT 5

US-10-156-761-3232

Sequence 3232, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

```

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 3232
LENGTH: 678
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(678)
US-10-156-761-3232

```

Alignment Scores:

Pred. No.:	9.57e-12	Length:	678
Score:	149.00	Matches:	50
Percent Similarity:	40.98%	Conservative:	34
Best Local Similarity:	24.39%	Mismatches:	79
Query Match:	17.31%	Indels:	42
DB:	9	Gaps:	4

US-09-674-779B-2 (1-172) x US-10-156-761-3232 (1-678)

```

QY 9 AlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLysSerThr 28
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 22 GCCACCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 29 SerGlnValMetValAlaProAsnAla-----ProThrGlyTyrThr 42
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 82 CTCACCTCAACCGTC---CCCTGAGCGGTGCGGAGCGAGTGAACCCACCGCGCGCGCG 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 43 GlyValIleTyrThrGlyValAlaProLeuValAsp----- 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 GCGACCGTGACCACTCGCCGCCACCCCGGAGGTGATCGGACGACACCGCGGTGAACTC 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 ----- 54
DB 199 GCGGAGGAGCGACCTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258
    -----AsnAspGluThrValLysAlaLeuAlaSerLysLeuProSer 68
QY 55 ----- 68
DB 259 CTCGCGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 69 LeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 319 GAGGTGCTTCCCGAGAGGACGCCGAGCTGAACCGCGAGCGCGCGCGCGCGCGCGCG 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 89 GluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThr 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 379 GCCATCGCGGAGAGATCAAGAGACGAAAGCCGCGCGCGCGCGCGCGCGCGCGCGCG 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 109 AspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaVal 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 439 GACAACCTGGGCTCGTACCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 129 ArgAsnTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleSerPhe 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 499 CACGACCAACTGCGCTCGCGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 148 GlyGluGluArgProIleAlaPheGlyThrAsnGluAlaIleTrpSerGlnAsnArg 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 559 AGCGAGGACTACCCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 168 AlaGluLeuSerTyr 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 619 GTGAGGTGCGCTTC 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

US-10-156-761-1/C

Sequence 1, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

```

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:

```



```

US-09-974-300-1902
; Sequence 1902, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2001-03-27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1902
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1902

```

```

Alignment Scores:
Pred. No.: 9.55e-07 Length: 561
Score: 113.50 Matches: 35
Percent Similarity: 44.728 Conservative: 20
Best Local Similarity: 28.46% Mismatches: 63
Query Match: 13.18% Indels: 5
DB: Gaps: 3

```

US-09-674-779B-2 (1-172) x US-09-974-300-1902 (1-561)

```

QY 53 ValAspAsnAspGluThrValIysAlaLeuAlaSerLysLeuProSerLeuValIyrPhe 72
Db 150 GTCATACGAGAGCTACAGACGAGCGCTCCCTTCATCCAGAGATATATCTTTTC 209
QY 73 AspPheAspSerAspGluIleLysProGlnAlaIleLeuAspGluIleAlaGln 92
Db 210 GATTCCGAAAGCGGAGATCCGCGACGAGACATCCGCTGCCAAGAGATATCCGAC 269
QY 93 PheLeuThrThrAsnGlnThrAlaArgValLeuAlaGlnIleThrAspGlu----- 110
Db 270 CTTCGCGATTCGAAACCCCGCCGCAATATCTTAATCCGCGCATACGAGCAATGTCGC 329
QY 111 --ArgGlySerArg--GluTyrAsnMetSerLeuGlyIleArgAlaValAlaVal 128
Db 330 ATTCGAATTCATTAATCAATGCGATTTAAGCGTGAATGCGGCTGCAATTC 389
QY 129 ArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerVal---GluIleIleSerPhe 147
Db 390 ATGGGGCTTTATATCGAAATCCAAAGCTTGACGCCCAAGATCTTCACGCGGAAAGGCTAC 449
QY 148 GlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaIleTyrSerGlnAsnArgArg 167
Db 450 GGGGAATTTAAACGATCGCTCAATAGACACCGAAGAGAGAGAGAGAGAGAGAGAGAGAG 509
QY 168 AlaGluLeu 170
Db 510 GTTGAATATC 518

```

RESULT 12

```

US-09-998-279-25
; Sequence 25, Application US/0998279
; Publication No. US20030083287A1
; GENERAL INFORMATION:
; APPLICANT: BURGESS, NICOLA A.
; APPLICANT: GARCIA, MIGUEL M.
; APPLICANT: KIRKE, DAVID F.
; APPLICANT: MEYERS, NICHOLAS L.
; TITLE OF INVENTION: gins
; FILE REFERENCE: GMS0081

```

```

; CURRENT APPLICATION NUMBER: US/09/998,279
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,288
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Porphyromonas gingivalis
US-09-998-279-25

```

```

Alignment Scores:
Pred. No.: 2.55e-05 Length: 1143
Score: 107.00 Matches: 40
Percent Similarity: 42.75% Conservative: 16
Best Local Similarity: 30.53% Mismatches: 49
Query Match: 12.43% Indels: 26
DB: Gaps: 6

```

US-09-674-779B-2 (1-172) x US-09-998-279-25 (1-1143)

```

QY 35 ProAsnAlaProThrGlyTyrThrGlyValIleLysThrGlyValAlaProLeuValAsp 54
Db 799 CCTACACAGCCTTAC---GTTACTCGTGTAGTC-----GTTGAC 834
QY 55 AsnAspGluThrValIysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPhe 74
Db 835 AAT-----GTGGTTACTCCGTAAT 855
QY 75 AspSerAspGluIleLysProGlnAlaIleLeuAspGluIleAlaGlnPheLeu 94
Db 856 AATATGTCGAAGATGATGCTTAATCAAGAAATCAATGTTCAATACAGCTGAATATGCG 915
QY 95 ThrThrAsnGlnThrAlaArgValLeuValAlaGlnIleThrAspGluArg--GlySer 113
Db 916 AAGACCAAC--AACGACCGATCAAGATGATGATGATGATGATGATGATGATGATGATGAT 972
QY 114 ArgGluTyrAsnMetSerLeuGlyIleArgAlaValAlaValAlaValAlaValAlaVal 133
Db 973 GCGGCTATATCAATGAAAGCTTTCAGAGCGCTGCGCAAAAGCGGTACCAAGATGCTTGA 1032
QY 134 GlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluIleArgProIle 153
Db 1033 AAGTATGATGTTTCTCGGATCGCATTAATGAATGAAGAGGCTCATCAGACCAATATC 1092
QY 154 AlaPheGlyThrAsnGluGluAlaIleTyrSerGln 164
Db 1093 -----TATGAAGAGAAAGCTTGAATCG 1116

```

RESULT 13

```

US-09-974-300-1889
; Sequence 1889, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2001-03-27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1889
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1889

```

Alignment Scores:

Pred. No.: 6,82e-05 Length: 669
 Score: 101.50 Matches: 34
 Percent Similarity: 40.918 Conservative: 20
 Best Local Similarity: 25.768 Mismatches: 63
 Query Match: 11.798 Indels: 15
 Gaps: 4

US-09-674-779b-2 (1-172) x US-09-974-300-1889 (1-669)

QY 53 ValAspAsnAspGluThrValIleuAlaLeuAlaSerLys----- 65
 Db 267 ATGGAAGAAACCGACGTCACAGTTGATGCGGCGGACGACGACGCGGCGCTCAT 326
 QY 66 -----LeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAla 83
 Db 327 CTCGCTCTTCAGAACGCGTTTGTTCATTCAGGAAGCGGATCTGAAGCAAGCT 386
 QY 84 AlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThraSngInThrAlaArgValIleu 103
 Db 387 CACCGCGTTTGCATTAATCGCCGCTCTTCATAATCGGTGCCAACCCG---ATTCGT 443
 QY 104 ValAlaGlyHisThrAspGluArgGlySerArgGluTyr-----AsnMetSer 119
 Db 444 GTGGAGGGCGCATCGGACGACGCGCCGATTCGACTTACCGCTTTCATCCATTCGGAG 503
 QY 120 LeuGlyGluArgAlaValAlaValArgAsnTyrLeuLeuGlyLys---GlyIleAsn 138
 Db 504 CTTTCTGCACAGAGGCGGACGACGCTGATCGGCTACTTACGTCGAAGAAAGCTCGAC 563
 QY 139 GlnAlaSerValGluIleIleSerPheGlyGluGlnArgProIleAlaPheGlyThrAsn 158
 Db 564 TCGGCGGCTTCTTTCGATCGGTTATGCGGATACAAACCGGTCAGGACACCCGACT 623
 QY 159 GluGlnAlaTyrPserGlnAsnArgAlaGluLeu 170
 Db 624 GAGAGCCATATGAGAGAAACAGCGCGCTCGAGATT 659

RESULT 14

US-10-010-160-11
 ; Sequence 11, Application US/10010160
 ; Publication No. US20030103999A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosey, Everett L.
 ; APPLICANT: Strugnell, Richard A.
 ; APPLICANT: Good, Robert T.
 ; APPLICANT: King, Kendall W.
 ; TITLE OF INVENTION: NOVEL THERAPEUTIC COMPOSITIONS FOR
 ; TREATING INFECTION BY LAWSONIA SPP.
 ; FILE REFERENCE: DAV110.001AUS
 ; CURRENT APPLICATION NUMBER: US/10/010.160
 ; CURRENT FILING DATE: 2001-11-09
 ; PRIOR APPLICATION NUMBER: AU PRI381
 ; PRIOR FILING DATE: 2000-11-10
 ; PRIOR APPLICATION NUMBER: US 60/249,596
 ; PRIOR FILING DATE: 2000-11-17
 ; NUMBER OF SEQ ID NOS: 68
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 717
 ; TYPE: DNA
 ; ORGANISM: Lawsonia intracellularis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(714)
 ; US-10-010-160-11

Alignment Scores:
 Pred. No.: 0.00109 Length: 717
 Score: 93.50 Matches: 36
 Percent Similarity: 42.218 Conservative: 29
 Best Local Similarity: 23.38% Mismatches: 70

Query Match: 10.86% Indels: 19
 Gaps: 6

US-09-674-779b-2 (1-172) x US-10-010-160-11 (1-717)

QY 24 AlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrGly 43
 Db 256 GCTAATATA---TTAAAAAAATGTTAATGGCTGATCTCATCCACAGTCAGCAAGCA 312
 QY 44 ValIleTyrGlyValAlaProLeuValAspAsnAspGluThrValIleValAla 63
 Db 313 ATAAGTCGTGACGATGTGGTGATTAATA-----CGTGAATTTCTAATTTC 360
 QY 64 SerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAla 83
 Db 361 ACGTTTTTCTGTTAGACGACACCTTCACCCGAAAGGAAAGATTATGGAGACTGT 420
 QY 84 AlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThraSngInThrAlaArgValIleu 103
 Db 421 TTAGCGCTTCCTCCGTAATATATATCTTTACCTT-----GTG 456
 QY 104 ValAlaGlyHisThrAsp-----GluArgGlySerArg---GluTyrAsnMet 118
 Db 457 ATACGTGGCATGCTGATATTGTGAATTAACAAAGGACGCGCTTTCCTCAACTCG 516
 QY 119 SerLeuGlyGluArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsn 138
 Db 517 GACCTTCAGAGGCGCTGCGACGACGACCTGCACAGATCTTGTAGAGCAGCGGATTAAG 576
 QY 139 GlnAlaSerValGluIleIleSerPheGlyGluGlnArgProIleAlaPheGlyThrAsn 158
 Db 577 GCTTCACGATTCGCTGCTGATGAGATGCAATTCACATTCACACCTCTAGAA---CCTAGTCT 633
 QY 159 GluGlnAlaTyrPserGlnAsnArgAlaGluLeuSerTyr 172
 Db 634 CCGAAGAGATGACAAAAATCGTGTAGATTCATTTT 675

RESULT 15

US-08-424-550b-80
 ; Sequence 80, Application US/08424550B
 ; Patent No. US20020119447A1
 ; GENERAL INFORMATION:
 ; APPLICANT: JOHN N. SIMONS
 ; APPLICANT: TAM J. PILOT-MATINS
 ; APPLICANT: GEORGE J. DAMSON
 ; APPLICANT: GEORGE G. SCHLAUDER
 ; APPLICANT: SURESH M. DESAI
 ; APPLICANT: THOMAS P. LEARY
 ; APPLICANT: ANTHONY SCOTT MURERHOF
 ; APPLICANT: JAMES C. ERKER
 ; APPLICANT: SHERI L. BUTIK
 ; APPLICANT: ISA K. MUSHAWAR
 ; TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS
 ; NUMBER OF SEQUENCES: 716
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
 ; STREET: 100 ABBOTT PARK ROAD
 ; CITY: ABBOTT PARK
 ; STATE: IL
 ; COUNTRY: USA
 ; ZIP: 60064-3500
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/424,550B
 ; FILING DATE:
 ; CLASSIFICATION: 435435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: FOREMSKI, PRISCILLA E.

SEQ ID NO 486
LENGTH: 4224
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
US-09-841-132-486

Alignment Scores:
Pred. No.: 0.819 Length: 4224
Score: 82.00 Matches: 42
Percent Similarity: 34.29% Conservative: 18
Best Local Similarity: 24.00% Mismatches: 81
Query Match: 9.52% Indels: 34
Gaps: 7

US-09-674-779b-2 (1-172) x US-09-841-132-486 (1-4224)

```
QY 15 SerValLeuThrPheMet-----ThrGlyCysAlaAsnLys 26
Db 1276 GCATCTCCTACCTTTATTCATTCAGAGTCCTGAACGCAATAAAGAGCACTATCGATTGCA 1335
QY 27 SerThrSerGluMetValAlaProAsnAlaProThrGlyThrGlyValIleThr 46
Db 1336 AACACCAAGAAAGTACGCTCCTGATGCTGACAGACAGTACGAGTGGTGGCGATCTAT 1395
QY 47 ThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeu 66
Db 1396 GCTACCAATGCTACTCTTACTGGAACGCTCC-----1428
QY 67 ProSerLeuValIlePheAspPheAspSerAspGluIleLysProGlnAlaAlaIle 86
Db 1429 -----CTGACCTTTCAGCGCAATCTGCTGCAACTTCAGAGGGCGCATC 1473
QY 87 LeuAspGluGlnAlaGlnPheLeuThrThrasnglnThrAlaArgValLeuValAlaGly 106
Db 1474 TATACAGAAATCGAAGATTACTCTTACAGGAAGTACGAGAACCGGACCTTCAGACACA 1533
QY 107 HisThrAspGluArgGlySerArgGluIle-----AsnMetSerLeu-----GlyGlu 122
Db 1534 AATACAGCAAGAACAGCGCGCCCTTATATCTTAAGGCAACACTCTGCTGTGTAT 1593
QY 123 ArgArgAlaValAlaValArgAsnTyrLeuGlnGlyLysGlyIleAsnGlnAlaSerVal 142
Db 1594 ACCAACCTGCTCTTTCAGGACCAACACTACGGGCGAGTATCTTCAGCAAAATCAA 1653
QY 143 Glu-----IleIleSerPheGlyGluGlnArgProIleAlaPheGlyThr 157
Db 1654 GAGGCTTCGGTGGGCGCATCTATCGTTCTTGAAGTCAGCATCTGTAGT-----ACT 1707
QY 158 AsnGluGluAlaTyr---SerGlnAsnArgArgAlaGluLeuSer 171
Db 1708 AAAAAAGAGCTGTGATGAGATTAAGAAACAAAGCTGAGTCTCTCT 1752
```

RESULT 23

US-09-841-132-171
Sequence 171, Application US/09841132
Patent No. US20020061848A1
GENERAL INFORMATION:
APPLICANT: Bhatia, Ajay
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Probst, Peter
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
FILE REFERENCE: 210121.469C8
CURRENT APPLICATION NUMBER: US/09/841.132
CURRENT FILING DATE: 2001-04-23
NUMBER OF SEQ ID NOS: 599
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 171
LENGTH: 2895
TYPE: DNA
ORGANISM: Chlamydia
US-09-841-132-171

Alignment Scores:
Pred. No.: 0.63 Length: 2895
Score: 81.00 Matches: 39
Percent Similarity: 37.58% Conservative: 23
Best Local Similarity: 23.64% Mismatches: 75
Query Match: 9.41% Indels: 28
Gaps: 7

US-09-674-779b-2 (1-172) x US-09-841-132-171 (1-2895)

```
QY 8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLysSer 27
Db 829 GCTTGTGAGGAGAGCAACATTAAGTAACTACTCCCTGATGATGATACAGCAATCGT--- 885
QY 28 ThrSerGlnValMetValAlaProAsnAlaProThrGlyThrGlyValIleThr 47
Db 886 ---GATAGATCTTTTGTGACATATACAAAAAATTATGGCGGAGCTATTACGCT 942
QY 48 GlyValAlaProLeuValAspAsnAspGluThr-----ValLysAlaLeuAlaSerLys 65
Db 943 CCTGTAGTACCTTGTGATGATGAGCTTACTCTTATTAACATATGCCAATAT 1002
QY 66 LeuProSerLeuValIlePheAspPheAspSerAspGluIleLysProGlnAlaAla 85
Db 1003 AAGGGGGCGGCTATCTATTAAGCAACGATAAC-----TCCAAAT 1047
QY 86 IleLeuAspGluGlnAlaGlnPheLeuThrThrasnglnThrAlaArgValLeuValAla 105
Db 1048 TCTGCCAGCCCGCATGATATTATTTAATGAATAATTTGACATATGATACAAATGA 1107
QY 106 GlyHisThrAspGluArgGlySerArgGluIleThrAsnMetSerLeuGlyGluArgAla 125
Db 1108 AATGTACACAGTACGCTACGTAATCTCT-----AGAGAAATATGA 1149
QY 126 ValAlaValArgAsn-----TyrLeuGlnGlyLysGlyIleAsnGln-----139
Db 1150 ATACAGTAGCAAGCTCTCTGTGAATTTCTATTAGACAGCAGGACTACCAAAATTTA 1209
QY 140 -----AlaSerValGluIleIleSerPheGlyGluGlnArgProIleAlaPheGly 156
Db 1210 ATTTTATGATCTTATTAAGATTAAGCAATGCAAGG-----GTCTGTGTGCTC 1257
QY 157 ThrAsnGluGluAla 161
Db 1258 TTCATATGAGAGCT 1272
```

RESULT 24

US-09-841-132-183
Sequence 183, Application US/09841132
Patent No. US20020061848A1
GENERAL INFORMATION:
APPLICANT: Bhatia, Ajay
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Probst, Peter
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
FILE REFERENCE: 210121.469C8
CURRENT APPLICATION NUMBER: US/09/841.132
CURRENT FILING DATE: 2001-04-23
NUMBER OF SEQ ID NOS: 599
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 183
LENGTH: 2934
TYPE: DNA
ORGANISM: Chlamydia
US-09-841-132-183

Alignment Scores:
Pred. No.: 0.643 Length: 2934
Score: 81.00 Matches: 39
Percent Similarity: 37.58% Conservative: 23
Best Local Similarity: 23.64% Mismatches: 75
Query Match: 9.41% Indels: 28

[illegible]

```

Db      574 AAGCTGACCAAGCAGCTGATGCTGATGCGGAGCG-----ACCGCGGACAG 621
Qy      46 TyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeu-----62
Db      622 TACACCTCGGTGGAGCACAAGAGAGACCTCACCAGCAAGGTCAACAGCAGTGGTGGACCGC 681
Qy      63 ---AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLysPro 81
Db      682 GCCCGCCGACAAGGTGGTGGACCGCGTCCGCGTGGACGGCGCGCGCGCGCGAGGCG 741
Qy      82 GlnAla-----AlaAlaIleLeuAspGluGlnAlaGlnPhe-----93
Db      742 CCGACGCTGAAGTCCGGCTCTTACAGGACCGTGGAGTGTGCCAGACCGCTGTATC 801
Qy      94 -----LeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGlu 110
Db      802 CGCGTGGACGTGAAGCGCGGCGCAGAGCTCCGCGCTCGGTGAGCTCTCCGCGACCGT 861
Qy      111 ArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsn 130
Db      862 CAGGTGAACCCGAGCTACGCGGGGTGTGTG-----CGGCGGTGACCGTGCACAC 912
Qy      131 -----TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db      913 CGTGAAGTCTGCGCGGTGAGCGCGCGGCGCAGCGCGCTGACGAGCTGTGCGACGGGT 972
Qy      149 GluGluArgProIleAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArgArgAla 168
Db      973 CTGCGCTATCCGAAGCCGAGAGCAGCAGCAGCAGCGACCGACCGCGACCGGTGTCTC 1032
Qy      169 GlnLeuSerTyr 172
Db      1033 CAGGTCTCCAC 1044

RESULT 27
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

```

Alignment Scores:
Pred. No.: 2.66e+05
Score: 80.00
Percent Similarity: 41.308
Best Local Similarity: 23.378
Query Match: 9.29%
Matches: 43
Conservative: 33
Mismatch: 84
Indels: 24
Gaps: 6

```

```

US-09-674-779b-2 (1-172) x US-10-156-761-1 (1-9025608)
Qy      6 GlnIleAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsn 25
Db      2445562 GAGATCGCGCGCCCAAGGCGCTCGGCGTCCACATGAGACACCTCGGTGGTCCGCAACAG 2445621
Qy      26 LysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIle 45
Db      2445622 AAGCTGAGCACAAGAGCTCAGCTCATCGCCAGAGCG-----ACCGCGGACAG 2445669
Qy      46 TyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeu-----62
Db      2445670 TACACCTCGGTGGAGCACAAGAGAGCTCCACCGCAAGGTCAACACCTGGTGGACCGC 2445729
Qy      63 ---AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLysPro 81
Db      2445730 GCCCGCCGACAAGGTGGTGGACCGCGTCCGCGTGGACGGCGCGCGCGCGAGGCG 2445789
Qy      82 GlnAla-----AlaAlaIleLeuAspGluGlnAlaGlnPhe-----93
Db      2445790 CCGACGCTGAAGTCCGGCTCTTACAGGACCGTGGAGTGTGCCAGACCGCTGTATC 2445849
Qy      94 -----LeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGlu 110
Db      2445850 CGCGTGGACGTGAAGCGCGGCGCAGAGCTCCGCGCTCGGTGAGCTCTCCGCGACCGT 2445909
Qy      111 ArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsn 130
Db      2445910 CAGGTGAACCCGAGCTACGCGGGGTGTGTG-----CGGCGGTGACCGTGCACAC 2445960
Qy      131 -----TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGly 148
Db      2445961 CGTGAAGTCTGCGCGGTGAGCGCGCGGCGCAGCGCGCTGACGAGCTGTGCGACGGGT 2446020
Qy      149 GluGluArgProIleAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArgArgAla 168
Db      2446021 CTGCGCTATCCGAAGCCGAGAGCAGCAGCAGCAGCGACCGCGACCGCGGTGTCTC 2446080
Qy      169 GlnLeuSerTyr 172
Db      2446081 CAGGTCTCCAC 2446092

RESULT 28
US-09-974-300-6221
; Sequence 6221, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Beika, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 2000-10-06
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: PASCSEQ for Windows Version 4.0
; SEQ ID NO 6221
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6221

```

```

Alignment Scores:
Pred. No.: 0.0437
Score: 78.50
Percent Similarity: 44.44%
Best Local Similarity: 25.56%
Query Match: 9.12%
Matches: 315
Conservative: 23
Mismatch: 17
Indels: 45
Gaps: 5

```


QY 21 ThrGlycSAlaAsnIysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
 Db 463 TGGTATGCTCGCTCTTGTCCGG -GAGGAGGCGGTCGGGGGAGCGGTCCTCCGAGAGA 405
 QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAsp----- 56
 Db 404 GCGCAGCGCGCGGCTCTCTCCGGGTGCGGAGACGTCTCGATGAGATGGCCGCGCGGG 345
 QY 57 ---GluThrValIysAlaLeuAlaSerIysLeuProSerLeuValTyrPheAspPheasp 75
 Db 344 GAAGCGGAGTCGACGCGGCGCTCTCCCTCGCTTTCAGACCGCTG----- 300
 QY 76 SerAspGluIleIysProGlnAlaAlaAlaIleLeuAspGluGln-----AlaGlnPhe 93
 Db 299 ---GATAGCCCTTCGCGCGACGCGAGCTTTCAGCCGCGCGCGCTGCTCTCCGATT 243
 QY 94 LeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySer 113
 Db 242 GCCGTACACAGCGGATATCTCGATGCTCTCCGCGTCCACACATCTTCGCGCGCGG 183
 QY 114 ArgGluTyrAsnMetSerLeuGlyIuArgArgAlaValAlaValArg-AsnIyrLeuLeu 133
 Db 182 CGATTTCGGAACGTCCGCTCTCGACCGAAGCTCGACCGCCGACCTGCGATTATTGT 123
 QY 133 uGlyIysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgProIle 153
 Db 122 TGGGGGTTGGATGAGGAAGGATCGATCGAT----- 92
 QY 153 eAlaPheGlyThrAsnGluGluAlaTyrSerGlnAsnArgArgAlaGluLeuSerTyr 172
 Db 91 -----CAGGAGAACCTTGTGTACTAGGGCGGAGCGAGAGGTGCGGTAT 47
 RESULT 31
 US-10-156-761-6208
 ; Sequence 6208, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMIKURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI

```

RESULT 35
US-10-156-761-6208
; Sequence 6208, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HASTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6208
; LENGTH: 2631
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2631)
US-10-156-761-6208

```

Predicted Scores:	
Pred. No.:	1,73
Score:	77.50
Percent Similarity:	34.55%
Best Local Similarity:	24.09%
Query Match:	9.00%
DB:	9
Length:	2633
Matches:	53
Conservative:	23
Mismatches:	87
Indels:	57

US-09-674-779B-2 (1-172) x US-10-156-761-6208 (1-2631)

QY 10 AlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLysSerThrSer 29
||| |||||||::: ||| :::
Db 802 GCGCGGGCGGCCTACAGAGCGCGTCTTCGCCGGCGGGGGCGCGGTGGTCTCC 861
QY 30 GluValMetValAlaProAsnAlaProThr-----GlyTyrrThrGlyValIleTyrrThr 47
||| ||||| :||| ||| |||||:
Db 862 GCGAGTCCCGTGCCGCAATCTCTCCCTACACGCTCCGGGTGTGGTCTGTGGACTGC 921
QY 48 GlyValAla-----ProLeuValAspAsnAspGluThrValLysAlaLeuAlaSer--- 64
|||:::|||| | ||| |||||||::: ||| ||||| |||:
Db 922 GGACTGCGCGGAGCACCATGTGGATCAGCGCGCGGCGGTGAGCGCCCTACACAGCTA 981
QY 65 -----LysLeuproSer 68
982 CGGCGCTCGCAGGGGGGGGGGGCGCACGGCGGGCGCGCGCGCGCGCGCGG 104
QY 69 LeuValTyrrPhe-----AspPheAspSerAsp 77
||| ||||| ||| ||| |||
Db 1042 ACGGTGAAGCTGCTGGGCGAGCGAGACGCCCTGTGCGCGGTTCCTCCCTCCC 110
QY 78 GluIleLysPro-----GlnAlaAlaAlaIleLeuAspGlu 89
||| ||||| ||| ||||| ||| |||
Db 1102 GAGATCAAAGGTGGCCGACCTGACAGCGCTTCCGTTGCGACGGGGGTGCTGGGCGATTCCG 116
QY 90 GlnAlaGlnPheLeuThr-----ThraSngInThrAlaArg 101
::: ||| ||| -----
Db 1162 GAGCGCTGCCGCTCGCTCTGTGACGCGCGCGGGCGGGGGGAGAGCGCGCGCGCGG 122
QY 102 ValLeuVal-----AlaGlyIsthrAspGluArgLysSer 113
|||::: ||| ||||| ||||| |||
Db 1222 TCGCTTCTCCGCCGATCCGCTCCGTGGATTCGCCGGGGCGGGCCACAGCGACGAGGGTG 1281
QY 114 ArgGluTyrrAsnMetSerLeuGlyLysuArgAlaValAlaValArgAsnTyrrLeuLen 133
|||::: ||| ||| |||||::: |||
Db 1282 CCGATGTCCCGGCTCGGCTGCACCCCGGACTGGCGGGGCTCTGCTGGACGGCGGCGCG 1341
QY 134 GlyLysGlyTyrLeuSnglAlaSerValGluIleIleSerPheGlyLysLysuArgProIle 153
||| :|||::: |||::: ||| ||||| |||
Db 1342 GAGGTGCGCGGCGGCGGGCGCGGAGGTGTCCGCTGTGACGAGGAGCGCCCGCGG 1401
QY 154 AlaPheGlyTyrrAsn--GluGluValATrPserGlnAsnTrgArgAlaGluLeuSerTyrr 172
::: ||| ::| ||||| ||||| ||||| |||
Db 1402 GAATACGCGGACGACTGGCGGCGCGCGTGGGGGTCTCCCGGCTGGGGGCGACCGCTAC 1461

RESULT 32
US-09-815-242-9978
Sequence 9978, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA, O11A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931

1

10

Db 37 GTGGAAGTGTGCGCTACGGGCTCCCGAGCTGACGTGGCTTGAATTGGAGGGGGCAG 96
Qy 79 ILeYsProGlnAlaAlaAlaLeuAspGluGlnAlaGlnPheLeu----- 94
Db 97 GTGAGAGCTCAGCGGCGAGCGCTTCTAGAAATGCTGAGCCGATTCAGGATTCAGCAAT 156
Qy 95 -----ThrThraGlnGlnThraAlaArgValLeuVal 104
Db 157 GTTGTTCATGATATATCAGAGAGATGATGACACTGATCAGAAATATATAGGGCTCCCTA 216
Qy 105 AlaGlyHisThrAspGluArgGlySerArgGlyIuTyraMetSer-----Leu 120
Db 217 GACCTGAAATTCACCAAGATTCGACATGGAATTAATTAATCAACAGAAAGATTTT 276
Qy 121 GlyIuArgArgAlaValAlaValArgAsnTyrlleuLeuGlyLysGlyIleAsnGlnAla 140
Db 277 CAGGAGCGCGCTGCTTATGCA-----GAGCAATG 306
Qy 141 SerValGluIleIleSerPheGlyGluIuArgProIleAlaPheGlyThrAsnGluGlu 160
Db 307 GTGGTGAGCTAAAGATATTATAGACAGAAAGATGTTCAACTGACAGAAAGATGAA 366
Qy 161 AlaTrpSerGlnAsnArgArgAla 168
Db 367 GCTCTACAGAGAGAGAGAAAGCT 390
RESULT 35
US-10-156-761-5966
; Sequence 5966, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5966
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(465)
US-10-156-761-5966
Alignment Scores:
Pred. No.: 0.219 Length: 465
Score: 75.50 Matches: 38
Percent Similarity: 38.65% Conserved: 25
Best Local Similarity: 23.31% Mismatches: 55
Query Match: 8.77% Indels: 45
DB: 9 Gaps: 7
US-09-674-779b-2 (1-172) x US-10-156-761-5966 (1-465)
Qy 7 ILeAlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLys 26
Db 25 GTGGCGGAGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 75
Qy 27 SerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrlleuValIleTyr 46
Db 76 AGCGGCTGCGAGAGAGCGGCGCGGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 126

Qy 47 ThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLys 66
Db 127 -----ACCCTGAGAGACTGCGCGCGAGGTG 153
Qy 67 ProSerLeuValTyrlleuPheAspPheAspSerAspGluIleLysProGlnAlaAlaAla 86
Db 154 -----AAGTGCAGAGCCGACATCCAGAGGAG 180
Qy 87 LeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGly 106
Db 181 GCCGACGAGATCCGCGAGGCGCTGTCAGAGACCTCCAGAAATTCATCTCCGCGAC 240
Qy 107 HisThrAspGluArgGlySerArgGlyIuTyraMetSerLeuGlyGluArgArgAlaVal 126
Db 241 TTGCCACACGACCGCGCGCGCGAGCTG-----CTCAGACAGCGCAAGAGCTAC 291
Qy 127 AlaValArgAsnTyrlleuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleSer 146
Db 292 GGC---GGTTCATACCTGCTCGGC-----GCCAAGTGGTGGCG 327
Qy 147 PheGlyGluGluArgProIleAlaPhe-----GlyThrAsnGluGlu 160
Db 328 GTGGGAGCAGAGAGGTGTCAAGCGCTGCGGAGACGCTCGCGGAGCATGTGAGAG 387
Qy 161 AlaTrpSer 163
Db 388 GGAAGTTCG 396
RESULT 36
US-09-815-242-4642
; Sequence 4642, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.01A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4642
; LENGTH: 2382
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4642
Alignment Scores:
Pred. No.: 2.88 Length: 2382
Score: 75.50 Matches: 38
Percent Similarity: 38.75% Conserved: 24


```

; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1335)
US-10-156-761-1454

Alignment Scores:
Pred. No.:      1 37          Length:      1335
Score:          75.00        Matches:       41
Percent Similarity: 36.46%    Conservative: 25
Best Local Similarity: 22.65% Mismatches:   45
Query Match:     8.71%       Indels:       70
DB:              9          Gaps:         9

US-09-674-779B-2 (1-172) x US-10-156-761-1454 (1-1335)
QY      29  SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleThrGly 48
Db      1104 GCTCGGGGTGGGGGTGCACCCAGATGCCGGCCCCGGCGTCACCCTGGAGAGTCAACCCCGGT 1045
               ||| ||| ||| ||| |||
QY      49  ValAlaProLeuValAspAsnSrgLutThrValIysAlaLeuAlaSerLysLeuPro-- 67
Db      1044 GTCCGCCCGCCGCCGT---GTGCATCAGACAGCGCGCGTGCCGCATCCCATCCGATGCCGTAC 988
               ||||| ||| ||| ||| |||
QY      68  -----Ser 68
Db      987  GACGTTCGTCAGTGCCTCTGAGCAGATCCGGAAGCGCCGAGATCGTGCAGCGCGCAG 928
QY      69  LeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAsp 88
Db      927 CTTCGGCACCTCCGCCCTGATCTGCACCGT---CAACCCCGCGGTCCGCCGTCTGCTGCAC 871
               ||| ||| ||| ||| |||
QY      89  GluGlnAlaGln-----PheLeuThrThraSn 97
Db      870 CAGTTCGCACAGCCGCGTGCAGCCCGGGCGCGCGCCGCGCAGCGCGCTGTGCGCGCTGCG 811
               ||| ||| ||| ||| |||
QY      98  Gln-----ThrAlaArgValLeuValAlaGlyHisThrAspGlu 110
Db      810 CAGGGGTGCGAGTAGTCAGTCCCGCACCTCTCCACAGCGCTCTGTGCGCGCGCTTGATGAT 751
               ||| ||| ||| ||| |||
QY      111 ArgGlySerArgLutTyrAsnMetSerLeuGlyLutArgAlaValAlaValArgAsn 130
Db      750 -----GTrAGCGCGCTGCG----- 736
               ||| ||| ||| ||| |||
QY      131 TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerVal----- 142
Db      735 CGCTGTCTCGGGGTGCGAGTGCAGACACGGGAGAACCGACCCCGCTGAACTGGATCAC 676
               ||||| ||| ||| ||| |||
QY      143 -----GluIleIleSerPhe----GlyLutGluArgProIleAlaPhe 155
Db      675 CGAATCTGTGCGCGCAGAGCTGTGCAATTGCGGGCGATCGCGAGCGCGTCTGCTGTC 616
               ||| ||| ||| ||| |||
QY      156 Gly 156
Db      615 GGC 613

RESULT 39
US-09-815-242-4085
; Sequence 4085, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A

```

QY 149 lucluarproleialaphelythr-asngluAlaIatrpsergluAsnarArgala 168
Db 758 AGAGCTAGATTAC---CTGGGACCCCGCCACGAGCGCTGCTCCGCTTCCCGGAGAGT 814

RESULT 40

US-09-938-956-5/C
; Sequence 5, Application US/09938956
; Patent No. US20020142408A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Siqun
; APPLICANT: Dicosimo, Deana J.
; APPLICANT: Kofas, Matheos
; APPLICANT: Odom, J. Martin
; TITLE OF INVENTION: Production of Monoterpene
; FILE REFERENCE: C11809 US NA
; CURRENT APPLICATION NUMBER: US/09/938, 956
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/229, 907
; PRIOR FILING DATE: 2000-09-0
; PRIOR APPLICATION NUMBER: 60/229, 858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 11575
; TYPE: DNA
; ORGANISM: Plasmid
US-09-938-956-5

Alignment Scores:

Pred. No.:	41.2	Length:	11575
Score:	75.00	Matches:	39
Percent Similarity:	37.89%	Conservative:	22
Best Local Similarity:	24.22%	Mismatches:	64
Query Match:	8.71%	Indels:	36
DB:	10	Gaps:	6

US-09-674-779b-2 (1-172) x US-09-938-956-5 (1-11575)

QY 10 AlaAlaAlaAlaSerValIleuThrPheMetThrGlyCysAlaAsnLysSerThrSer 29
Db 10095 GCCGGGCGACGCTTCACGCGCTCAATTCCTGACGGGCTGT----- 10054
QY 30 GluValMetValAlaProAsnAlaProThrGlyThrGlyValIle--TyrThrGly 48
Db 10053 -----TTCCTTTGCGGAGTCGCAAGGCGAGCCGCCCTTACGCCGGAGGCT 10000
QY 49 ValAlaProLeuValAsp-----AsnAspGluThrValLysAla----- 61
Db 9999 CTCACCCCGCTGCTTCGTCGCGGCGCGGCGCATGACCGTCGTCGCCGCCCTGATG 9940
QY 62 -----LeuAlaSerLysLeuProSerLeuValTyrPheAspPhe 74
Db 9939 GCGGCTTCTTCATCATGCACTGCGACAGGTCGCCGCCGCCCTTGGGTCAATTTC 9880
QY 75 AspSerAspGluIleLysProGluAlaAlaIleLeuAspGluGlnAlaGlnPheLeu 94
Db 9879 GCGGAGATCGCTTCACATGCGGACCGACGACGATCGGATTCGCTGCGCATTTGGC 9820
QY 95 ThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArg 114
Db 9819 ATTCTGCATTCACCTGCCAGCAATGATCACCGGCCCTGTAGCCCGCG----- 9769
QY 115 GluTyrAsnMetSerLeuGluArgArgAlaValAlaValArgAsnTyrLeuLeuGly 134
Db 9768 -----CTCGCGAAGCGGCGACTC-----ATGCTCGGA 9739
QY 135 LysGlyIleAsnGlnAlaSerValGluIleSerPheGlyGluGluArgProIleAla 154
Db 9738 ATGATTCGCGGACGACAGGCTACATCTGCTTCCCTCGCAGACAGGAGATGATGGCG 9679
QY 155 Phe 155

Db 9678 TTC 9676

Search completed: July 6, 2003, 15:04:07
Job time : 3796 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:20 ; Search time 1609 Seconds
(without alignments)
3111.052 Million cell updates/sec

Title: US-09-674-779b-2
Sequence: 1 MMHIIQIAAANAALSVLTQM.....IAFGNNEAMNSQNRRLAST 172
Perfect score: 861

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV-xlp
-O=/cgn2.1/USPO.spool/US09674779/runat_06072003.121613.9801/app_query.fasta.1.327
-DB=genhmbl -GMMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=100
-DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -ECGN 1.1319 -erunat_06072003.121613.9801 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESOURCY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONCLIG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Genhmbl:.*
1: gb_ba:.*
2: gb_htg:.*
3: gb_in:.*
4: gb_cm:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vl:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_cm:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_un:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	861	100.0	519	6 AX011034	AX011034 Sequence
2	861	100.0	92407	6 AX067461	AX067461 Sequence
3	855	99.3	519	6 AX011036	AX011036 Sequence
4	850	98.7	519	6 AX011038	AX011038 Sequence
5	844	98.0	519	6 AX011040	AX011040 Sequence
6	316	36.7	7577	1 PEPALI	X74218 Pseudomons
7	307	35.7	1048	1 PAOPRTGN	X50191 P. aeruginos
8	306	35.5	13263	1 AE004530	AE004530 Pseudomon
9	289	33.6	20941	1 AE008730	AE008730 Salmonell
10	289	33.6	294050	1 AL627268	AL627268 Salmonell
11	287.5	33.4	6967	1 ECH29785	AJ29785 Brwlinia c
12	286.5	33.3	10029	1 AE005987	AE005987 Caulobact
13	285.5	33.3	212050	1 AL646050	AL646050 Ralstonia
14	285.5	33.2	7713	1 ECPAL	X05123 E. coli pal
15	285.5	33.2	1531	1 ECXECGC	X65796 E. coli exc
16	285.5	33.2	10653	1 AE000177	AE000177 Escherich
17	285.5	33.2	11670	1 AE005252	AE005252 Escherich
18	285.5	33.2	16419	1 D90713	D90713 Escherichia
19	285.5	33.2	297816	1 AP002553	AP002553 Escherich
20	283	32.9	11138	1 AE013906	AE013906 Yersinia
21	283	32.9	210050	1 AE014146	AJ14146 Yersinia
22	276.5	32.1	12393	1 AE012414	AE012414 Xanthomon
23	272.5	31.6	11713	1 AE011958	AE011958 Xanthomon
24	269.5	31.3	349116	1 AP003003	AP003003 Mesorhizo
25	265.5	30.8	10833	1 AE004009	AE004009 Xylella f
26	264	30.7	462	6 AR089419	AR089419 Sequence
27	264	30.7	737	1 HPA15KLP	M1878 H. influenza
28	264	30.7	737	6 I06718	I06718 Sequence 3
29	264	30.7	737	6 I06718	I06718 Sequence 2
30	264	30.7	737	6 I06718	I06718 Sequence 1
31	264	30.7	867	6 HEAOMPP6	M19391 H. influenza
32	264	30.7	13620	1 U32722	U32722 Sequence 1
33	262.5	30.5	340900	1 SME591791	U32722 Haemophilus
34	262	30.4	10906	1 AE009302	AE009302 Agrobacte
35	262	30.4	11049	1 AE008312	AE008312 Agrobacte
36	259.5	30.1	10643	1 AE012837	AE012837 Chlorobiu
37	259.5	30.1	11914	1 AE004259	AE004259 Vibrio ch
38	258	30.0	393	1 PMU271673	AJ271673 Pasteurel
39	258	30.0	453	1 AE006136	U16489 Pasteurella
40	258	30.0	11061	1 AE006136	AE006136 Pasteurel
41	257.5	29.9	237523	1 RPPX04	AJ235273 Pasteurel
42	254	29.5	845	1 BRUOMP16A	L27996 Brucella ab
43	254	29.5	5937	1 AF358662	AF358662 Brucella
44	254	29.5	10099	1 AE009476	AE009476 Brucella
45	253.5	29.4	1213	1 APDNPALA	X89009 A. pleuropne
46	246	28.6	669	1 HDU42466	U42466 Haemophilus
47	245	28.3	1540	1 LPLIPO	X60543 L. pneumophi
48	243.5	28.0	10359	1 AE008667	AE008667 Rickettsi
49	241	28.0	635	1 BAOUNKN	L47642 Bartonella
50	215.5	25.0	540	6 AX055138	AX055138 Sequence

BASE COUNT	139 a	108 c	122 g	150 t	
ORIGIN	/organism="Moraxella catarrhalis" /db_xref="taxon:480"				
Alignment Scores:					
Pred. No.:	1,96e-82	Length:		519	
Score:	861.00	Matches:		172	
Percent Similarity:	100.008	Conservative:		0	
Best local Similarity:	100.008	Mismatches:		0	
Query Match:	100.00%	Indels:		0	
DB:	6	Gaps:		0	
US-09-674-779B-2 (1-172) x AX011034 (1-519)					
Oy	1 MetMetLeuHisiIlegInIIeAlaAlaAlaAlaAlaLeuSerValIeuThrPheMet	20			
Db	1 ATGATGTACATATTCAAATGCGCGCGCGCGCGCTTATTCGGTACTAAGCTTTTANG	60			
Oy	21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly	40			
Db	61 ACAGGCTGTGCCAATAAATCATCAACAGTCAGTATGTGCTGCTCAATGCACCCACAGGT	120			
Oy	41 TyrThrGlyValIleIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys	60			
Db	121 TACATGTGGGTTATCTATACGTGCTGTGCACCTTTGGTATGATGATGAGACCGTTAAG	180			
Oy	61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys	80			
Db	181 GCTTCGGCAAGCAAGCTCCACCGTTGGTTTATTTTGACTTGATTCGATGACATTTAA	240			
Oy	81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla	100			
Db	241 CCGCAGCTGCGTCCATCTTTACAGCAACAGCAACATTTTAAACCCAAATCAACAGCT	300			
Oy	101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu	120			
Db	301 CGTGTGTTGGTTCAGGTCTATCCGATGCGCTGGTAGTCTGAGATATAATATGCACTG	360			
Oy	121 GlyGluArgArgAlaValAlaValaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla	140			
Db	361 GGGGAACGCGTGGCTGGCGGAGTACCCAACTATTTGGCTTGTAAGCATTAATCAACCC	420			
Oy	141 SerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu	160			
Db	421 AGCGTTGAGTTATCAAGTTTGGTGAAGAAGCCCATTCGATTTGGCACAAATGACAGA	480			
Oy	161 AlaTrpSerGlnAsnArgAlaGluLeuSerTyr	172			
Db	481 GCATGCTCACAAAATCGTCTGCTGAACGTCTTAT	516			

	RESULT 2			PAT 24-JAN-2001
AX067461/c	DNA	linear		
LOCUS	92407 bp			
DEFINITION	Sequence 36 from Patent WO0078968.			
ACCESSION	AX067461			
VERSION	AX067461.1 GI:12545081			
KEYWORDS	.			
SOURCE	Moraxella catarrhalis.			
ORGANISM	Moraxella catarrhalis Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;			
REFERENCE	Moraxella . 1 (bases 1 to 92407)			
AUTHORS	Lagace,R.E., Peterson,C. and Berg,K.L. Nucleotide sequences of moraxella catarrhalis genome Patent: WO_0078968-A 36 28 -DEC-2000; Incyte Genomics, Inc. (US) 			
JOURNAL	Location/Qualifiers			
FEATURES	I..92407			
source	/organism="Moraxella catarrhalis" /db_xref="taxon:480"			
BASE COUNT	26788 a 17581 c 20150 g 27888 t			
ORIGIN				

Alignment Scores:

Pred. No.: 1.51e-79
Score: 861.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%

Length: 92407
Matches: 172
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-09-674-779b-2 (1-172) x AX067461 (1-92407)

QY 1 MettelleuhtstlleglnllleallalaalalaaleuSerValleuthrPhem 20
DB 26057 ATGATGTTACATATTCATAATTCGCCGCCGCCGCTTATTCGCTACTTATG 25998
QY 21 ThrGlyCysAlaansylSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
DB 25997 ACAGGCTGTCACATATTAATACACAGTATGATGCTCTTAATGACCCACAGT 25938
QY 41 TyThrGlyValleThrThrGlyValAlaProLeuValAspAsnAspGluThrVallys 60
DB 25937 TACACTGGGGTATCTACTGCTGTTGCACCTTGTGATATGATGACCGTTAG 25878
QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLeys 80
DB 25877 GCTCTGCAAGCAAGCTACCCAGTTGGTTATTTTGACTTGTGATTCGATGAGATTAA 25818
QY 81 ProGlnAlaAlaAlaAlaLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
DB 25817 CCCCAAGCTGCTGCCATCTTACGACACACACACATTTTACCACCAATCAACAGCT 25758
QY 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
DB 25757 CGGTTTGGTTGCGAGGTATACCGATAGCGGTGCTGCTGATATATATGTCACCTG 25698
QY 121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
DB 25697 GGGGAACGCGCTGCGGTGGCTAGCAACATTTGCTGTGTAAGCATTAATCAAGCC 25638
QY 141 SerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu 160
DB 25637 ACCGTTGAGATTATAGTTTGGTGAGAACGCCCTATCCGATTGGCACAAATGAGAA 25578
QY 161 AlaTyrSerGlnAsnArgArgAlaGluLeuSerTyr 172
DB 25577 GCATGTGCACAAATCGTCTGCTGACTGCTTAT 25542

RESULT 3
AX011036

LOCUS AX011036 519 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO957277.
ACCESSION AX011036
VERSION AX011036.1 GI:997680

KEYWORDS

SOURCE

ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.

REFERENCE 1 (bases 1 to 519)

AUTHORS Ruelle,J.L.

TITLE Basb019 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses

JOURNAL Patent: WO 957277-A 3 11-NOV-1999;

RUELE JEAN LOUIS (BE); SMITHKLINE BEECHAM BIOLOG (BE)

FEATURES

source

1. 519
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"

BASE COUNT 139 a 108 c 121 g 151 t

ORIGIN

Alignment Scores:

Pred. No.: 8.5e-82 Length: 519

Score: 855.00

Percent Similarity: 99.42%
Best Local Similarity: 99.42%
Query Match: 99.30%

Matches: 171
Conservative: 0
Mismatch: 1
Indels: 0
Gaps: 0

US-09-674-779b-2 (1-172) x AX011036 (1-519)

QY 1 MettelleuhtstlleglnllleallalaalalaaleuSerValleuthrPhem 20
DB 1 ATGATGTTACATATTCATAATTCGCCGCCGCCGCTTATTCGCTACTTATG 60
QY 21 ThrGlyCysAlaansylSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
DB 61 ACAGGCTGTCACATATTAATACACAGTATGATGCTCTTAATGACCCACAGT 120
QY 41 TyThrGlyValleThrThrGlyValAlaProLeuValAspAsnAspGluThrVallys 60
DB 121 TACACTGGGGTATCTACTGCTGTTGCACCTTGTGATATGATGACCGTTAG 180
QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLeys 80
DB 181 GCTCTGCAAGCAAGCTACCCAGTTGGTTATTTTGACTTGTGATTCGATGAGATTAA 240
QY 81 ProGlnAlaAlaAlaAlaLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
DB 241 CCCCAAGCTGCTGCCATCTTACGACACACACACATTTTACCACCAATCAACAGCT 300
QY 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
DB 301 CGGTTTGGTTGCGAGGTATACCGATAGCGGTGCTGCTGATATATATGTCACCTG 360
QY 121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
DB 361 GGGGAACGCGCTGCGGTGGCTAGCAACATTTGCTGTGTAAGCATTAATCAAGCC 420
QY 141 SerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu 160
DB 421 ACCGTTGAGATTATAGTTTGGTGAGAACGCCCTATCCGATTGGCACAAATGAGAA 480
QY 161 AlaTyrSerGlnAsnArgArgAlaGluLeuSerTyr 172
DB 481 GCATGTGCACAAATCGTCTGCTGACTGCTTAT 516

RESULT 4
AX011038

LOCUS AX011038 519 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 5 from Patent WO957277.
ACCESSION AX011038
VERSION AX011038.1 GI:997681

KEYWORDS

SOURCE

ORGANISM Moraxella catarrhalis.
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.

REFERENCE 1 (bases 1 to 519)

AUTHORS Ruelle,J.L.

TITLE Basb019 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses

JOURNAL Patent: WO 957277-A 5 11-NOV-1999;

RUELE JEAN LOUIS (BE); SMITHKLINE BEECHAM BIOLOG (BE)

FEATURES

source

1. 519
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"

BASE COUNT 140 a 108 c 120 g 151 t

ORIGIN

Alignment Scores:

Pred. No.: 2.88e-81 Length: 519
Score: 850.00 Matches: 169
Percent Similarity: 98.84% Conservative: 1
Best Local Similarity: 98.26% Mismatches: 2


```

PSFMWPKDFIGPAGLLAARFLIDSRDTEITDRLSGNSDAFVFRCHSIMNCVSCP
KGLNTPRAIGHIKSMILGRSA"
gene
3811..6733
/ gene="sua"
/ note="STM0736"
3811..3819
/ gene="sua"
/ note="putative -35 signal for sua";
RegulonDB: STMTH004698"
-10 signal
3838..3846
/ gene="sua"
/ note="putative -10 signal for sua";
RegulonDB: STMTH004698"
RBS
3921..3926
/ gene="sua"
/ note="putative RBS for sua; RegulonDB: STMSIH001046"
3932..6733
/ gene="sua"
/ EC_number="1.2.4.2"
/ note="similar to E. coli 2-oxoglutarate dehydrogenase
(decarboxylase component) (AAC73820.1); Blastp hit to
AAC73820.1 (93% aa), 94% identity in aa 1 - 933"
/ codon_start=1
/ transl_table=11
/ product="2-oxoglutarate dehydrogenase decarboxylase
component"
/ db_xref="GI:16419246"
/ protein_id="AAL19680.1"
/ translation="MONSALKAMLDSSYLSGNSQWIEQLYEDFLTDPSVDANMRLT
FOGLPGTVKPDOLSHKSTREYFRQALAGSHSTISDPNTNKKVQLQILNAYFR
GHQANLIDPLGIMKOEYRADIDPSFHLDEADPEFTNDSFSGKEMKLELIDAL
KOTYCGPILGAEYMHITSTREKRWIOORIEGSRAPASDEKREFLNELTAEGELRIG
AEFGAKRFSLEGGDALIPMLKEVRRRAGNSGTEVYIGMAHGRNLNVLNVGKSO
DLDFEPAKHEHLEGTGVKYHMGFSSDIETEGVLALAFNPSHLETYSPPVMSO
RALRLRDEPSSNKVLPITIGDAVVGQVQDPLTMSKARGIEVGTVRIYINNOY
GFTTNPIDARSTPYCDIGKVOAPIFHNVAADPEAVAEVTRIALDFNTERDVI
DLYCYRRHGHNEADPSATOPLOMOKIKHPTPKRIADKLEADKVALDDEATENVL
VRDALDAGECVYKEMRPMNMHSTFSPPLNHEMDEAYPNKEMKRLQELAKRSTPE
AIENOSRVAKIYCDROMAAGEKLEFMDGAEMLAYATVDEGIFRVLGSDSGRGTFE
HHRHAYHNQNSTTTPROHITHSGOGKRVMSDISEAVLAFEYGAATAPRLLTIL
EAQDFDPAQAVYIDQFTISGEOKKMGCLVNLPHGVEGOGPEHSSARLELYTL
CAENOVQVCPSTPAQVYHMLRQALRMKRPVVMSPKSLRRPLAVSTLDELANS
FQPAIGEIDELDPKAVKRYVMCSKRVYDLEQRKNDQKVALVRIEQLPPPHKAV

```

```

Alignment Scores:
Pred. No.: 1 03e-19 Length: 20941
Score: 289.00 Matches: 67
Percent Similarity: 54.95% Conservative: 33
Best Local Similarity: 36.81% Mismatches: 64
Query Match: 33.57% Indels: 18
DB: 1 Gaps: 3

```

```

US-09-674-779b-2 (1-172) x AE008730 (1-20941)
QY 1 MetwelfeuhisilecgnilealaalaalaalaaleuSerValleuthrPhemet 20
DB 19478 ATGCAACTGAACAAGGCTGAGGCGCTGATGATGCGCTGTTATGCAATCGCG 19537
QY 21 ThrcglysalaanslysserthrSerGlnValMetValAlaProAsnAlaProthrly 40
DB 19538 GCATGTTCTTCCACAGAACGCCACGCAAT-----GACGGT 19573
QY 41 TyrThrGlyValIleYrThrGlyValAlaProLeuValAsp-----54
DB 19574 AGCAAGCGGTATGCTGACGCGCGGCGCATGTATGACGCTAACGCAAGCGCAAC 19633
QY 55 -----AsnAspGluThrValLysAlaLeuAlaSerLysLeuPro-----SerLeuVal 70
DB 19634 ATGTCATCTGTAAGGCAAGCGCTGACATCAGACGCTGACGACGACGACACATCGTT 19693
QY 71 TyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGluGln 90
DB 19694 TACTTCGATCTCGACAGATGATATCCGTTGACTTCGCGGCAATGCTGATGCGGAC 19753

```

```

QY 91 AlaglnPheLeuThrThrasnGlnIhrAlaArgValLeuValAlaGlyHisThrAspGlu 110
DB 19754 GCTAACTCTCTGTTGACCAACCGCTCTTCAACAGTACACCGTACAGAGTACAGCGGAGAA 19813
QY 111 ArgGlySerArgGluThrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsn 130
DB 19814 CGCGGATACCGGAGTACCAACATCTCCCTCGGTGAGCGCTGCTAACGCCGTTAAATG 19873
QY 131 TyrLeuLeuGlyLysGlyLysAlaGlnAlaSerValGluIleIleSerPheGlyGluGlu 150
DB 19874 TACCTGCGGCTAAAGCGCTTCCGCTGACACAGATCTCCATCTTCTTACGTTAAACA 19933
QY 151 ArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgAlaGluLeu 170
DB 19934 AAACCTCGCTACTGGGCCACGACGAGGCGGCTTACGTTAAGAACCGCTGCTACTG 19993
QY 171 SerTyr 172
DB 19994 GTTTAC 19999
RESULT 10
AL627268 294050 bp DNA linear BCT 06-JUN-2002
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome; segment 4/20.
ACCESSION AL627268 AL513382
VERSION AL627268.1 GI:16501953
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi.
Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
1 (bases 1 to 294050)
Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,
Wain, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T. G.,
Sebatbia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Comerton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N.,
Farrar, J., Felwell, T., Hamlin, N., Hague, A., Hien, T. T., Holroyd, S.,
Jagels, K., Krogan, A., Larsen, T. S., Leather, S., Moule, S., O'Gaora, P.,
Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S., and Barrall, B. G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
Nature 413 (6858), 848-852 (2001)
21534947
MEDLINE
PUBMED
2 (bases 1 to 294050)
11677608
REFERENCE
Parkhill, J.
Direct Submission
Submitted (25-Oct-2001) submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES
source
1..294050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/strain="CT18"
/db_xref="taxon:90370"
249..1343
/ gene="STY0759"
249..1343
/ gene="STY0759"
249..1343
/ gene="STY0759"
/EC_number="5.4.99.9"
/ note="Similar to Klebsiella pneumoniae probable
udp-galactopyranose mutase rfbD SW:GLP_KLEPN (Q48485)
(384 aa) fasta scores: E(0, 76.7% id in 315 aa, and to
Escherichia coli udp-galactopyranose mutase SW:GLF_ECOLI

```

```

() (367 aa) fasta scores: E(): 0, 38.3% id in 339 aa
Paralogue of E. coli yefE (GUF_ECOLI); Fasta hit to
GUF_ECOLI (367 aa), 39% identity in 340 aa overlap"
/codon_start=1
/transl_table=11
/product="UDP-galactopyranose mutase"
/protein_id="CAD05181.1"
/db_xref="GI:16501954"
/db_xref="SPRMBL:0828D7"
/translation="MSANRYTKIKNQKRLIIVAGISGAVIAEQGHVNIIDOR
SHIGSNAYDARDEHTGIMVYVYGHPIPHDNEIVNMYVNTAMPEYINKVAVTQ
VPSPIVHTINHPFVACSDPARKLLQKCDSTILEPQNEOQALRIGELAEAF
PKGTTIKWGLHSPALPASVAKRIPVAFENDYDIPFNHKGFOGIPKFGYIQMWYSIEHE
NIAVELCRSTQEMRTNVDHVFSGALADAFYSCQYRLRYLIDPEFKITICSDYQCA
VMNYSIDTYEYRTTEHKYSPWREHNASICYEYSRECEADDPYIPVRADKMDLL
NKYLSRAKKEKILLSLVYLAIVLISWTIS"
1426..2319
/gene="STY0760"
1426..2319
/gene="STY0760"
/translation="Similar to Klebsiella pneumoniae rfbE rfbE
TR:Q48482 (EMBL:L41518) (297 aa) fasta scores: E(): 0,
57.9% id in 297 aa, and to Shigella flexneri
dtdp-rhamnosyl transferase RfbF rfbE SW:RFBF_SHFL
(P37782) (296 aa) fasta scores: E(): 5.6e-08, 24.4% id in
303 aa
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase"
/protein_id="CAD05182.1"
/db_xref="GI:16501955"
/db_xref="SPRMBL:0828D6"
/translation="MKCALIVTFNRLKRSVRETVKAGFSSIVIVNGSSDQTR
WSSLSPEGITILLKNDLGGAGFKVGSQYCSYNADWFEYDDDAYEPINLIKFS
SLDTSVSGMIFRVDYIGRSCRMNLEPIRVSPVEFTIYVAMREPSVRRQVND
VQTVSFVGMIDIRKVLNHLNDIDELFYVDPEFGYKIVLSQKIRISPEIKFID
ISIHGRICPEHKKVYICNLLKLLKPLVRPIISVTLVLSKYATALLPWOKKER
YLFIWQGITLHKGISGKYH"
1435..1914
/gene="STY0760"
/translation="Pfam match to entry PF00535 Glycos transf.2.
glycosyl transferases, score 12.60, E-value 0.013"
2332..3465
/gene="STY0761"
2332..3465
/gene="STY0761"
/translation="Similar to Klebsiella pneumoniae
galactosyltransferase wbo or rfbF TR:Q48487 (EMBL:L31762)
(377 aa) fasta scores: E(): 0, 69.1% id in 375 aa, and to
Serratia marcescens galactosyltransferase wbo or rfbF
TR:Q54481 (EMBL:L34167) (380 aa) fasta scores: E(): 0,
55.2% id in 375 aa"
/codon_start=1
/transl_table=11
/product="galactosyltransferase"
/protein_id="CAD05183.1"
/db_xref="GI:16501956"
/db_xref="SPRMBL:08X721"
/translation="MKRLCTRVNSDFDLHWTERATAADAGYEIHTISFTGEI
KKFTLIGICNLSVLAQSENMVEFFRAFLNARKIIEINPDILHCTLKPCIGVS
ARKRSNPVVISVGLGRFLYVPMRILALVLYVKYTAGNKGVPFIEHKKDRK
ISRLVGIDHKTIVIEGAGINPDYIKRSIEKKHPIVLPASRMKSGKGLDEAKK
ILROKNIHFVNAGIIEADDDKDAIPLELHHMHNGILNMLGRSSNYVELIOKSNIV
ALPSIPEGVPRLLLEASSVGRACIADPGCCSLIHNHNGILVKSNAQELAVELLE
YLLKPOIRLEMGNGKRRKVELFSLIVLTKLQIKDTIGC"
2872..3390
/gene="STY0761"
/translation="Pfam match to entry PF00534 Glycos transf.1.
glycosyl transferases group 1, score 152.20, E-value
8.9e-42"
3594..4364
/gene="STY0762"
/pseudo
misc_feature
/translation="conserved hypothetical protein (pseudogene)"
7138..7971
/gene="STY0767"
CDS
3594..4364
/gene="STY0762"
/translation="Similar to Mycobacterium tuberculosis hypothetical
32.3 kDa protein cy13d12.17 rv3783 or mtcy13d12.17
TR:P72049 (EMBL:Z80343) (280 aa) fasta scores: E():
1.4e-26, 42.2% id in 187 aa, and to Yersinia
enterocolitica O-antigen export system permease protein
RfbD rfbD SW:RFB_DYEREN (Q56902) (259 aa) fasta scores:
E(): 9.3e-19, 29.4% id in 187 aa. This CDS contains a stop
codon at 188. The sequence has been checked and is
believed to be correct"
/pseudo
/codon_start=1
/transl_table=11
/product="putative polysaccharide export ABC transporter
permease protein (pseudogene)"
3603..4154
/gene="STY0762"
/translation="Pfam match to entry PF01061 ABC2 membrane, ABC-2
type transporter, score 21.40, E-value 5.2e-07"
/pseudo
4368..5078
/gene="STY0764"
4368..5078
/gene="STY0764"
/translation="Similar to Mycobacterium tuberculosis hypothetical
30.0 kDa protein cy13d12.15 rv3781 or mtcy13d12.15
TR:P72047 (EMBL:Z80343) (273 aa) fasta scores: E(): 0,
43.8% id in 233 aa, and to Yersinia enterocolitica
O-antigen export system ATP-binding protein RfbE rfbE
SW:RFB_EYEREN (Q56903) (239 aa) fasta scores: E(): 0,
45.3% id in 225 aa"
/codon_start=1
/transl_table=11
/product="putative polysaccharide export ABC transporter
ATP-binding protein"
/protein_id="CAD05185.1"
/db_xref="GI:16501957"
/db_xref="SPRMBL:0828D5"
/translation="MKISCKNVGVILPFIENSRSRPFKTFLOAAGGRIGSSNGTIE
VALKRIDFTTEGNRLALIGHNSGKTTLLRVLAGAVKPSGKKEGICVTSIDPM
NMGDELIGLENIKRLGLFLGSKNEIKNTEIDYERSEICDPKTIYRVYSSGMVLR
LGFSTAIINPEILLDEMSVGDSDPRKRAEMNLNFSKAGIMWATHDDELAHSV
CNKFTRLHSESLVSKGR"
4539..5051
/gene="STY0764"
/translation="Pfam match to entry PF00005 ABC tran. ABC
transporter, score 110.90, E-value 2.5e-29"
4560..4583
/gene="STY0764"
/translation="Pfam match to entry PF00005 ABC tran. ABC
transporter, score 110.90, E-value 2.5e-29"
5172..6945
/gene="STY0765"
5172..6945
/gene="STY0765"
/pseudo
5172..6945
/gene="STY0765"
/translation="Similar to Mycobacterium tuberculosis hypothetical
71.5 kDa protein rv3808c or mtcv026.13C TR:Q53585
(EMBL:AL022076) (637 aa) fasta scores: E(): 0.00015, 26.6%
id in 467 aa, and to Mycobacterium tuberculosis
hypothetical 33.9 kDa protein cy13d12.16 rv3782 or
mtcy13d12.16 TR:P72048 (EMBL:Z80343) (304 aa) fasta
scores: E(): 0.16, 25.6% id in 215 aa, and to Klebsiella
pneumoniae RfbE rfbE TR:Q48482 (EMBL:L41518) (297 aa)
fasta scores: E(): 0.3, 23.6% id in 233 aa. This CDS
contains at least one stop codon and frameshift. The
sequence has been checked and is believed to be correct"
/pseudo
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein (pseudogene)"
7138..7971
/gene="STY0767"
gene
3594..4364
/gene="STY0762"
/pseudo
misc_feature
/translation="conserved hypothetical protein (pseudogene)"
7138..7971
/gene="STY0767"
CDS
3594..4364
/gene="STY0762"
/translation="Similar to Mycobacterium tuberculosis hypothetical
32.3 kDa protein cy13d12.17 rv3783 or mtcy13d12.17
TR:P72049 (EMBL:Z80343) (280 aa) fasta scores: E():
1.4e-26, 42.2% id in 187 aa, and to Yersinia
enterocolitica O-antigen export system permease protein
RfbD rfbD SW:RFB_DYEREN (Q56902) (259 aa) fasta scores:
E(): 9.3e-19, 29.4% id in 187 aa. This CDS contains a stop
codon at 188. The sequence has been checked and is
believed to be correct"
/pseudo
/codon_start=1
/transl_table=11
/product="putative polysaccharide export ABC transporter
permease protein (pseudogene)"
3603..4154
/gene="STY0762"
/translation="Pfam match to entry PF01061 ABC2 membrane, ABC-2
type transporter, score 21.40, E-value 5.2e-07"
/pseudo
4368..5078
/gene="STY0764"
4368..5078
/gene="STY0764"
/translation="Similar to Mycobacterium tuberculosis hypothetical
30.0 kDa protein cy13d12.15 rv3781 or mtcy13d12.15
TR:P72047 (EMBL:Z80343) (273 aa) fasta scores: E(): 0,
43.8% id in 233 aa, and to Yersinia enterocolitica
O-antigen export system ATP-binding protein RfbE rfbE
SW:RFB_EYEREN (Q56903) (239 aa) fasta scores: E(): 0,
45.3% id in 225 aa"
/codon_start=1
/transl_table=11
/product="putative polysaccharide export ABC transporter
ATP-binding protein"
/protein_id="CAD05185.1"
/db_xref="GI:16501957"
/db_xref="SPRMBL:0828D5"
/translation="MKISCKNVGVILPFIENSRSRPFKTFLOAAGGRIGSSNGTIE
VALKRIDFTTEGNRLALIGHNSGKTTLLRVLAGAVKPSGKKEGICVTSIDPM
NMGDELIGLENIKRLGLFLGSKNEIKNTEIDYERSEICDPKTIYRVYSSGMVLR
LGFSTAIINPEILLDEMSVGDSDPRKRAEMNLNFSKAGIMWATHDDELAHSV
CNKFTRLHSESLVSKGR"
4539..5051
/gene="STY0764"
/translation="Pfam match to entry PF00005 ABC tran. ABC
transporter, score 110.90, E-value 2.5e-29"
4560..4583
/gene="STY0764"
/translation="Pfam match to entry PF00005 ABC tran. ABC
transporter, score 110.90, E-value 2.5e-29"
5172..6945
/gene="STY0765"
5172..6945
/gene="STY0765"
/pseudo
5172..6945
/gene="STY0765"
/translation="Similar to Mycobacterium tuberculosis hypothetical
71.5 kDa protein rv3808c or mtcv026.13C TR:Q53585
(EMBL:AL022076) (637 aa) fasta scores: E(): 0.00015, 26.6%
id in 467 aa, and to Mycobacterium tuberculosis
hypothetical 33.9 kDa protein cy13d12.16 rv3782 or
mtcy13d12.16 TR:P72048 (EMBL:Z80343) (304 aa) fasta
scores: E(): 0.16, 25.6% id in 215 aa, and to Klebsiella
pneumoniae RfbE rfbE TR:Q48482 (EMBL:L41518) (297 aa)
fasta scores: E(): 0.3, 23.6% id in 233 aa. This CDS
contains at least one stop codon and frameshift. The
sequence has been checked and is believed to be correct"
/pseudo
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein (pseudogene)"
7138..7971
/gene="STY0767"

```


QY 119 SerLeuGlyGluArgAlaValAlaValArgSerTyrLeuLeuGlyLysGlyTlleAsn 138
 :::
 Db 5576 GCGCTGGGCGACGTCGTCACACGCTGCAGATGACGAGGCGACGAGCGTGTCT 5635
 QY 139 GlnAlaSerValGluIleLeuSerPheGlyGluArgProIleAlaPheGlyThrAsn 158
 :::
 Db 5636 GCCGATCAGATCTCCCTGGTTCCTTACGGTTAAGAGAACTGCCGCTTCGGTCATGAC 5695
 QY 159 GluGluAlaTrpSerGlnAsnArgAlaValGluLeuSerTyr 172
 :::
 Db 5696 GAAGCTGATGATGTCACAAACCGCTGCGCTTGGTATAC 5737
 RESULT 12
 AE005987 10029 bp DNA linear BCT 12-JUN-2002
 LOCUS Caulobacter crescentus CB15 section 313 of 359 of the complete
 DEFINITION genome.
 ACCESSION AE005987
 VERSION AE005987.1 GI:13424910
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 1 (bases 1 to 10029)
 Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E.,
 Eisen, J., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R.,
 Potocka, I., Nelson, M.C., Newton, A., Stephens, C., Phadke, N.D.,
 Ely, B., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
 Haft, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H.,
 Shetty, J., Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
 Emoliaveva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and
 Fraser, C.M.
 Complete genome sequence of *Caulobacter crescentus*
 Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
 2 (bases 1 to 10029)
 Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J.,
 Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I.,
 Nelson, M.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B.,
 Laub, M.T., Deboy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
 Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
 Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J.,
 Emoliaveva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
 Fraser, C.M.
 TITLE Direct Submission
 JOURNAL Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1..10029
 /organism="Caulobacter crescentus CB15"
 /strain="CB15"
 /db_xref="taxon:190650"
 complement(79..1305)
 /gene="CC3227"
 complement(79..1305)
 /gene="CC3227"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK25189.1"
 /db_xref="GI:13424911"
 /translation="MRIDVRAALDRRLDRLDPAASRAPLAVGSGGDSILFLKTVLDMAR
 PLDRVLAIVYDHOPOSTAMTAEVAKAKALGALGLTWDOKRTGLPAAARA
 RRLATATAREAGARVILIGHATASDLAEGVRAEGSSVSPRAMAPSPWPGRLP
 VRLPLITTRAEIRDALTRBETVLDPANLDIYARARAAGALPLPLVRESPP
 GVFEIDPAGARLPRDVPYPAHLAALCASAREPPKCDRLARLVETLRSGARFTTL
 AGARIEDAVLICRDAGEAARGGLARDLDPGACGWDGKWEVYVAKRTLAYVALRG
 RTSSLPAGRARLSAIPAVRSLPVLISLDGDAPGELVLDLQLEADGARVRSVLP
 DREKAVGLDDESVT"
 complement(1296..2150)
 gene
 CDS
 1296..2150
 /gene="CC3228"
 complement(1296..2150)
 /gene="CC3228"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK25190.1"
 /db_xref="GI:13424912"
 /translation="MTLKSALLAVLFASTAAVAACPLPPADPLDRAKRYERNEK
 VYRELRSTVPGCRPKRVYVQPAFTDQALALNERVSDLEOTFLRLNGQESTNPEL
 TKAQKATDQKARADALEQRLAIEKTLADLQAAATQAAPLADGPPAPADAPA
 VAKQAKDLILAGDPAANEQAFALVNNYPSARTPPARVWIGETQVRRAYVDAAGN
 YIGIRGWPQTSWAPDAILKLARSWALRKTTECRLLDELAKRYRPAARQVNSRAS
 TRFAAKCA"
 complement(2228..2794)
 /gene="CC3229"
 complement(2228..2794)
 /gene="CC3229"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=11
 /product="Ompa family protein"
 /protein_id="AAK25191.1"
 /db_xref="GI:13424913"
 /translation="WRKMMSPFDTORARLALNGIAASIAACASRRKPPQVTPPPAO
 POPPTPTPPAPPVPTVSGILGSGVDFVYVNIQDRTYFDDEVIRADQPIYAGA
 QWLRYSVIRIEGNADRGREYNALGARRANAVADFLIAGVSSARIEITISRK
 ERIPIDSSSEEMAKNRNGRTAITDGAR"
 complement(2991..4289)
 /gene="CC3230"
 complement(2991..4289)
 /gene="CC3230"
 /note="similar to GB:M28232 SP:P19935 PID:148020 GB:000096
 PID:1651332; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="to1B protein"
 /protein_id="AAK25192.1"
 /db_xref="GI:13424914"
 /translation="MRRLALLMVAALLSGAFYPPANAOIEVDINAGAVKELPLAIP
 VFGSTRGAEIAIVITGNLERGLQPLNVAAGVPKLDADVQVQPPFPGSTGNOAL
 NGCVTVGADGSDVDRLMDVFSQDLGLIOPSSPPDMWRRAHISDAVERLTGK
 GYDTRAAFYAESGPKLRIKRLAIDMDGAPQVLTGSLYVMRPSYSQELTTH
 ALRPGSTYILNLETSROETRYKFGVAGFAFSDGSVAFSEVERGNSNDIYVMD
 RSRATPRITDPADIDTSPSPGKIVFNSDGGCAQIYVNTDGGSGVRRISYGGGR
 YTPVMSPRDPTAFTKQCGEPHIGVMRAGDGDDELITTSYLDGSPTMANGRVLMF
 FRGPGSGNPRMLVTDTIGRLRPAATSGSASDPANSPILD"
 complement(4319..5131)
 /gene="CC3231"
 complement(4319..5131)
 /gene="CC3231"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK25193.1"
 /db_xref="GI:13424915"
 /translation="MSARREOTLSPALGSIALHGVAAVIAFGLPWKOSKPTITGES
 VPYITVINGNPNVPAEALAEQTAOTPEPTPATQPPAPPPAPPPAPAPAPKPT
 PAKPTPKPTPAKPTPOKADNDNPFASLEIAKTKQKATGPANAPKGPAPAEFSVS
 PARPMGATGSLAALGRLOGEVODRNNPCEVEGGSNNVNVVYVIGGGRVQOPE
 SPGTSSPDVVKASDRAIRALFPAASFALVPSDLTGOKIALNFAKAKOCSR"
 complement(5139..5657)
 /gene="CC3232"
 complement(5139..5657)
 /gene="CC3232"
 /note="similar to PID:1103861 SP:P50599; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="Exbd/To1R family protein"
 /protein_id="AAK25194.1"
 complement(1296..2150)
 gene
 CDS


```

/db_xref="GI:13424916"/
/translacion="MAMSSNDPAFATGCGRRGRRRRRCALSEINVTPLVLMVLIRL  
IMTSPILITACYPDELEPTNDAALONDERPITSIKRGQIFVEETELPEFNLAAR  
GA1AGAGDHCGLFEVRADGGATAYVAOYWAGSNAGFNKINLITEGGPSGAPRAD  
QEGAGEDLPAP" complement(5662. .6360)  
/gene="CC3233"  
complement(5662. .6360)  
/note="identified by match to protein family HMM"  
/codon_start=1  
/transl_table=11  
/product="MotA/TolQ/ExbB proton channel family protein"  
/protein_id="AAK25195.1"  
/db_xref="GI:13424917"  
/translation="MDAAAPAAPNSFFPALFMQADMYKSVYTGILASLSGWAVIDLK  
LFEPALNRANRPREDIQRIDTOIARETTVEEGLSIAVAIASPIGLEFWYGIMAFNP  
MSERDAGLRILARRIBRIIDTOIARETTVEEGLSIAVAIASPIGLEFWYGIMAFNP  
DLNAISKTSTSIAYAVASIAELEFPALIGLINALPAYIVAINFSFDACKYAGRLSGFAD  
DISTAIRORLERY" complement(6475. .6918)  
/gene="CC3234"  
complement(6475. .6918)  
/note="identified by Glimmer2; putative"  
/codon_start=1  
/transl_table=11  
/product="conserved hypothetical protein"  
/protein_id="AAK25196.1"  
/db_xref="GI:13424918"  
/translation="MIETPTAGEVERHQQLPVRIYEEDTPFTGLIYHANYLTRLEGR  
SDFRANASHTELEAKODCTGFAYRMELDFRKRARIDDALVTRLEPRAGCVLRHLRQ  
TTRIGDDVLEAEATYAVACISLSGRPRRPTDMKLKSPWLIIPDS"  
complement(6915. .7781)  
/gene="CC3235"  
complement(6915. .7781)  
/note="identified by Glimmer2; putative"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein"  
/protein_id="AAK25197.1"  
/db_xref="GI:13424919"  
/translation="MAASSRLIRLLTAGIANGVANSIAAISAGAPRNALLIYOALA  
AGIATAGILELVGFMRWARWRMLAACGIALVLTLMGVSLLEGARKRVVALGPOLH  
TASIVLPFAAPAFAARCEGRAPRALATALLTLVOHDASISLAWALLAAALALERP  
ROAIIPACVAGIAGILLVGMVEDDTIPAPVYEGLDQDFEAPNAPVYAGVAGLMTSLP  
APBLULVARPERPERRAALLAALMTGWLAGNFGNYPAPVLGYGAPLWGLSIGLV  
LGDSVDQDPAP" complement(7771. .8811)  
/gene="CC3236"  
complement(7771. .8811)  
/gene="CC3236"  
/note="identified by match to TIGR protein family HMM"  
/codon_0635"  
/tcdon_start=1  
/transl_table=11  
/product="Holliad junction DNA helicase RuvB"  
/protein_id="AAK25198.1"  
/db_xref="GI:13424920"  
/translation="MTAVISGPHOGDIAPADRARPOTTALEFGVEDQAKCNLFVIE  
AAKGAGMALDHLLLEGGPEKTKTGTIOAVALELGVNFRATSGPYLNKPGDIAAILTNL  
RANVDFIDEIRHLSNNPEILYIPAMEDHVLDLVYIGESPASRSIRIDLAPTIIAAATP  
RAGMATPLRPFRCPIPLETETPRELSHVILQHAARKKGAPLTDGDADIELAKRARGP  
FVAGGLLARVDAFDRDADCDIDRKAAMALARLEVESGSDIDRLRYRAMIENYGS  
GPGVCETIATAIEARDRDVEVIEPLYMOQGFIQRTPRGMACKGAYLHLGTPPAP  
GEQAQGLFLFDSG"  
complement(8808. .9425)

```

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

Best Local Similarity:	38.248	Mismatches:	50
Query Match:	33.288	Indels:	21
DB:	1	Gaps:	4

Y	1	LveAlaValAlaAlaAlaAlaAlaAlaLeuSerValIleThrPhmeCthrGlycysAlaAsnLys	26
D	2740	GTCGCCCTGGCCGCCCGCTTGCCTGCCG-----GCCGCCCTTCGGGT	2695
Y	27	SerThrSerInValMetValAlaProAsnAla-----ProThrgly	40
D	2698	CCGAAGGCCACAGCCCGTCACGCCGCCGCCCGCTCAACCGCACCGACCACCCACCGACGCT	2639
Y	41	TyrThr-----GlyValIleIleYrThrGlyValAlaProLeuValAspAsnSp	56
D	2638	TACACGCCCGCGCGCTCCGCCGCCGCCGGGACCAACGCGATCCGCGCC-----	2594
Y	57	GltThrValIlysaIalaLeuAlaSerIylsIeuProSerLeuValTyrPhasPheAspSer	76
D	2593	GGTTCCGTGCAGACTTTCGTCTCATATATCGCCGACCGGGCTCTATTGTGACACCGACGAA	2534
Y	77	AspGIllleLysProGlnAlaAlaIleLeuAspGlunAlaGlnPheLeuThr	96
D	2533	TACGTCATTCGGCGCGCACCGCCCAACGGCGTGCGCGGCTCAAGCCCAATGGCTAACCCG	2474
Y	97	AsnGlnThrAlaValIleuValIlaGlylshTrAspGluValArglySerAglutyr	116
D	2473	TATTGCAGGCGTGCATCCGATCCGATCTCAAGACCAACGCCGACGACCGACCGGATAC	2414
Y	117	AsnMetSerLeuglyGluArgArGaIaValAlaValArgasnTyrlauLeuglyLySgly	136
D	2413	AACCTGGCGCTCGGGGCCGCCGCCCAACGCCGCGCGCGACTTCGTATCGCCACAGGC	2354
Y	137	IleaSnGlnAlaSerValGluIleIleSerPheGlyGluIuaArgProIleAlaPhegly	156
D	2353	GYTTGCAGCGGCGCGCATTCGATCAACAACGATCTCTTCGGCAAGAAGACCCCATTCGATCGCGGT	2294
Y	157	ThraSnGluGluAlaITPSPserGlnAsnArg	166
D	2293	TCGACGGAAGAAGCCTGGGCCCAAGAACCCG	2264
RESULT 13			
AL646060		212050 bp	DNA linear BCF 07-DEC-2001
LOCUS	Ralstonia solanacearum GM11000 chromosome, complete sequence;		
DEFINITION	segment 4/19,		
ACCESSION	AL646060 AL646052		
VERSION	AL646060.1 GI:17427566		
KEYWORDS	.		
SOURCE	Ralstonia solanacearum.		
ORGANISM	Ralstonia solanacearum.		
	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;		
REFERENCE	1 (bases 1 to 212050)		
AUTHORS	Salanoubet,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,		
	Attali,C.M., Billault,A., Brotier,P., Camus,J.C., Cattolico,L.,		
	Chandler,M., Choisme,N., Claudet-Renaud,C., Cunac,S., Demange,N.,		
	Gaspil,C., Lavie,M., Moisan,A., Robert,C., Sautin,W., Schlex,T.,		
	Siquier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,		
	Weissenbach,J. and Boucher,C.A.		
JOURNAL	Genome sequencing of the plant pathogen Ralstonia solanacearum		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 212050)		
TITLE	Boucher,C.A.		
JOURNAL	Submitted (05-DEC-2001) GenomeScope and CNRS UMR-8030 2 rue Cavoat		

COMMENT
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Bouchet@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1. 212050

FEATURES

source

gene

/organism="Rastonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
59. .757
/gene="RSC0556"
/note="RS04909"
59. .757

CDS

/gene="RSC0556"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD14086.1"
/db_xref="GI:17427567"
/translation="MALTSAPFWFALGSIILNIVISGNNAVIALARNPFRORO
ALFWGSAATIALRIALVLAIVKLLALPYLKTIGAVLLVYIGIKLSEADEAGSRP
HDGLMPAIRILLIADIVSLDNVAVAAAEKPGPDVALLVGLGSIPIIIGST
LVALMRLPIVITLAAALIGYLAGDMLVTPVDAGMPQAFYADVAIGCALIVV
IVGMMGRSLRQA"
951. .1472
/gene="pila1"
/note="RSC0557; RS04908"
951. .1472

gene

/gene="pila1"
/function="structural elements: cell exterior; surface
structures"
/note="Product confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TYPE 4 FIMBRIAL PILIN SIGNAL PEPTIDE
PROTEIN"
/protein_id="CAD14087.1"
/db_xref="GI:17427568"
/translation="MASKSRRRRGQGFLLIELMIVAILGILALAIAPAYDYTI
RSRYVGLALAGAKALYENANAKOSLSIGASLPSKNSALINDSSIGELAVVY
APAVTEAGANTIVITPYSSSVFQNLQAGSTPPAMVLMVCAAGKSPVNTVQRTT
ATLPPKAYSECR"
1739. .2245
/gene="pila"
/note="RSC0558; RS04907"
1739. .2245

CDS

/gene="pila"
/function="structural elements: cell exterior; surface
structures"
/note="Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="TYPE 4 FIMBRIAL PILIN SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD14088.1"
/db_xref="GI:17427569"
/translation="MKSRLNKRQKFTLLIELMIVAILGILALAIAPAYDYTV
ARVEGSLAAOKALYSENAANAKOSLSIGSSVFPPTKNSANLTIAGCTVPGQITV
TYTTAAGGTIALVPTNAGTALPVSSAPSPIMMTCYQNKQAASVAPSGTMSLAA
TYTTAAGGTIALVPTNAGTALPVSSAPSPIMMTCYQNKQAASVAPSGTMSLAA

CDS

gene

2351. 4060
/gene="RSC0559"
/note="RS04906"
2351. 4060
/gene="RSC0559"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD14089.1"
/db_xref="GI:17427570"
/translation="MLNFWYLAIVACNSVPLVAVSHSYPITFYSEFVAALIGWYALA
GVIGSTWHSKAGLPVYVAPLALIGVILVQVLYVNPINPPTFAIVELIGYAVGL
GACRDVPGVLESIAVAVITIGLLVLAIECHLPRVSGLPINMISTMPDGRMMGN
LNQPNHATVYLAEGLAACLEIGSTRRRVAPLALIALIGMALTYRSMSHLYVY
GVAGLAGAEEGARGARWIRAGVPVIGLAVYOLCNWLVAVANVLMHLDLPISDEL
OOSVGLRVFLMKHAMHMLAHPLMGSGCDVAMNOYVOTDVLGVHESMNAHNLVDL
LAKVYVGLLAVMLPELGLVHAAMRRMTPALAFYAVILTVASHSMLEYPLHYFL
LPFAVAGVYDPRKSRVLSPTNAVITIGVAVLIGRMWIDYQSEVERLYSLDGP
VELQYRSQGLLVPYIGLSTANNAGMTAFAPIMAAVEHQAVOYFPGSGTVQRMAL
ALAFQCKTDEALTVYRLHNOYVIDIAGDSKLLITVCTKKLGLATCARLAKENLVV
GVD"

CDS

gene

complement(4289. .4765)
/gene="moac"
/note="RSC0560; RS04905"
complement(4289. .4765)
/gene="moac"
/function="small molecule metabolism; biosynthesis of
cofactors, carriers; molybdopterin"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MOLYBDENUM COFACTOR BIOSYNTHESIS
PROTEIN C"
/protein_id="CAD14090.1"
/db_xref="GI:17427571"
/translation="WPHLHFDTRAGQAHMVDGDKAVTHRVAVAGLIMLPTEFALV
RDGTAKRGDVGIAIRIAIOGAKRTSLDPLCHRIALKYANDFEIDEDADRYRQTV
AETHQGTGVEMERALTAVOGLITTYDKCAVDKRWYIGDVLMEKKGSGSEWVAQ"
4859. .6544
/gene="RSC0561"
/note="RS04904"
4859. .6544
/gene="RSC0561"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD14091.1"
/db_xref="GI:17427572"
/translation="MTLSFRPRARKVTAVALMTWLAAPSPLAOPASVAPAPAGA
AAAEVQSNLNSVOIGROSPYLOQDATITAPYEIPDMGDPSTALSPMEBRIGD
RWQRQIRDPDVPYDPLSDYINDIGLITELARQHYAGTSAFPELAVRDPGIN
AFALPGYIGVNTVPLVATESSELASVIGHEIGVLDRIHARGLDKSGESMTALAS
ILLAGLAATKSGDAOLAVGQAAVAVSGLAFSGAGREADRVGFTLLTGAGTNPFG
MPFFRRLRVSSIADTGVAVPYARTHPLTGBRIADMEDRARGLPHPRQOPRPAFGFA

CDS

gene

4859. .6544
/gene="RSC0561"
/note="RS04904"
4859. .6544
/gene="RSC0561"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEICAL SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD14091.1"
/db_xref="GI:17427572"
/translation="MTLSFRPRARKVTAVALMTWLAAPSPLAOPASVAPAPAGA
AAAEVQSNLNSVOIGROSPYLOQDATITAPYEIPDMGDPSTALSPMEBRIGD
RWQRQIRDPDVPYDPLSDYINDIGLITELARQHYAGTSAFPELAVRDPGIN
AFALPGYIGVNTVPLVATESSELASVIGHEIGVLDRIHARGLDKSGESMTALAS
ILLAGLAATKSGDAOLAVGQAAVAVSGLAFSGAGREADRVGFTLLTGAGTNPFG
MPFFRRLRVSSIADTGVAVPYARTHPLTGBRIADMEDRARGLPHPRQOPRPAFGFA

CDS

gene
/protein_id="AAG55065.1"
/db_xref="GI:12513662"
/translation="MNCSEIRLKSTEMHLLMAKTTSSQCKLEVDHNYLNTLNQVFNK
LTLMTLYDTH"
complement(454..876)
/gene="20897"
complement(454..876)
/gene="20897"
/function="orf; unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55066.1"
/db_xref="GI:12513663"
/translation="MIPQISENTLPVATVASEFTYDQESDILYASCCSLKDAY
LSRCLASHPTLAKNIYETENLSDOKILHDCRSTNEIKTFEGQOREGLIETSDT
VFSILSLIDAIKSFQHOIPVNSVARELLIYDYD"
complement(958..1902)
/gene="20898"
complement(958..1902)
/function="orf; unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55067.1"
/db_xref="GI:12513664"
/translation="MSPDNIGSAVNRSPGSPNSDLTKDESKNTSPYOVYIKNTAI
LDPRIVAGIDLTLYTQRETFESQOLIESALOTLHVMTLNLERNRRESEYTLAE
YNDTSRTIVAKNEKPGCYGGLIODEDDNLKIKTFEARNVVAQMSDF
EFMFESYSAQONLEPHLLPAPVPLKTEPTEQEDVLLREFKMDPFKLISDE
DICEFVERDIPSPYRFEKFTMLALVDTLQTEANKIL"
complement(2077..2118)
/gene="20899"
complement(2077..2118)
/function="orf; unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG55068.1"
/db_xref="GI:12513665"
/translation="MSTDRKPVMLFPH"
2705..4276
/gene="cyda"
/note="20900"
2705..4276
/function="enzyme; Energy metabolism, carbon: Electron
transport"
/note="Residues 1 to 523 of 523 are 100.00 pct identical
to residues 1 to 523 of 523 from Escherichia coli K-12
strain MG1655: B0733"
/codon_start=1
/transl_table=11
/product="cytochrome d terminal oxidase, polypeptide
subunit I"
/protein_id="AAG55069.1"
/db_xref="GI:12513666"
/translation="MMLDIVELSRQPLTMYRLEVPILTGNAPLLATMEVYVLS
GKQIKDMTKRPGKLFIGNPAGVATGLTMEFQGTWSSYSHVGDIFGAPLAEI
NFEIRMEVSESELVNLNPAQVKEFVHTVASGYTAMFLLGISAWMYLGRQPAEK
NSFAIAASFMAAVLSYVLGDESGYEMGDVOKTALAIAEMETOPAPAAFTLFGIP
DOEETNKFAIOIPALGIIATRSVDTPVIGLKRKLQHEERIRNGKATSLLEQKRS
GSTDAVDQNSMKKDGGLLKRTTPNADTETAOIQATKDSIPRAAPLYEAFR
IMVAGFLLAIALFMSVYRNIRIGKKMLLRALYGIPLPWIAEAGFVAEYERQ
PWALGEVPTAVANSLAGDLIFSMVLIGLYTLFLVAFLLFMFARLGPSSLKGT
RYHPEQSSITTPAR"
4292..5431
/gene="cydb"
/note="20901"

CDS
4292..5431
/gene="cydb"
/function="enzyme; Energy metabolism, carbon: Electron
transport"
/note="Residues 1 to 379 of 379 are 100.00 pct identical
to residues 1 to 379 of 379 from Escherichia coli K-12
strain MG1655: B0734"
/codon_start=1
/transl_table=11
/product="cytochrome d terminal oxidase polypeptide
subunit II"
/protein_id="AAG55070.1"
/db_xref="GI:12513667"
/translation="MIDEVERFLTWLLNGVLLIGFVATDGPDMGMLTRELGRNDT
ERRMIINSLAHMGNQVWLITAGALPAPMPVYAAAFSGFYAMILVLALEFRPV
GEDRSKTEELRRMRMDMGIFISFEPPLVIGVAFGNLQGVFPVDEYLRYYGN
FPQILNPGLAGVSVGMITQGATYLOKRTVGEHLRTRATAQVAALVTLCFALA
GVWVYIGIDYVAKSTMDHYAASNPDKREVRAGAVLVNFTPLMALPALGVYLP
LTLITLARMDAKAWAFVSSLLACILITAGIAMFVPPSSPMNNAISLMDATSSQ
LTLNVTWAVAVLPILILYATMCYKMFGRITKEDIERNTHSLY"
5559..5852
/gene="ybgE"
/note="20903"
5559..5852
/gene="ybgE"
/function="orf; unknown function"
/note="Residues 1 to 97 of 97 are 100.00 pct identical to
residues 1 to 97 of 97 from Escherichia coli K-12 strain
MG1655: B0735"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55071.1"
/db_xref="GI:12513668"
/translation="MSKITATLYAVVMDKRPDLALSFVALLIAGCMFDPSPRAKTS
ELEHMGILIMAVACAGVHGVGFQKVLMOGIFCPDLADIVLIGLIFPF"
6002..6406
/gene="ybgC"
/note="20904"
6002..6406
/gene="ybgC"
/function="orf; unknown function"
/note="Residues 1 to 134 of 134 are 100.00 pct identical
to residues 1 to 134 of 134 from Escherichia coli K-12
strain MG1655: B0736"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55072.1"
/db_xref="GI:12513669"
/translation="MNTLFRPVRYIEDTDAGGVYHASYAFERARTEMLRHH
FSGQALMERVAFYRKMTVEYAPARIDMLLETQETTSKRGTSLVFTQRYVAENT
LLNEAEVLVCCYDPLMKRPLAPSVIAEFQ"
6403..7095
/gene="tolQ"
/note="20905"
6403..7095
/gene="tolQ"
/function="membrane; Colicin-related functions"
/note="Residues 1 to 230 of 230 are 100.00 pct identical
to residues 1 to 230 of 230 from Escherichia coli K-12
strain MG1655: B0737"
/codon_start=1
/transl_table=11
/product="inner membrane protein, membrane-spanning,
maintains integrity of cell envelope; tolerance to group A
colicins"
/protein_id="AAG55073.1"
/db_xref="GI:12513670"
/translation="MTDMNILLDLFKASLLVKLIMLIIIGFSIASAIIORTRLA
AAREAEAFEDKFMGIELSRLOYSGKRNLLSGQIYSGFKEVRLHRANSHAP
AVVEGASRARIRISNRELETHPIPLGVGSISPYIGLFTGVGIMHAFIALGAVK
QATLQWAPAIARLITATIGLFAAIPAVAAVNRILNRQVNLDELNYDNMFETAILH

gene
1168 . 3801
/gene="ybgp"
1168 . 3801
/note="ORF_ID:0178#1
similar to SwissProt Accession Number P54746"
/codon_start=1
/transl_table=11
/product="Hypothetical protein in hrsa 3' region ."
/protein_id="BA35398.1"
/db_xref="GI:4062319"
/translation="MKASVRYHTPHMHMDREWFTTEESRILLVNNMEILCREOD
NEKYVYLDGOTALEDEYPAVPEKPKRQYVAGLIIIGPWYQTDITTSASBIT
RLTMGMDCARGEPMKIGTLPDSFGMSGLPHIYNGEITRTMRCGSEHGNDK
TEFLMOSDSGEVTAQVPLGAYIGKYLPADENGILKRLDYSFDLEKASVKEILLP
NGHDMPLQONIPEVMDKREIYQKRFVREPEVEKLEAQRDLATLKGFEIDGK
YMRVHRTIGSTRMDIKIAHAREIKIYNLEPLATLATLKGFEYHGLLEKMKELK
NHAHDSIGCCSDVHREIYARELEADMDNLTIRYMKRIDNMFQSDADKLVLENL
MPMRREVINTVYLRASQFNLRDROCPYITIRAREIDPELIDRQIVHGNDFP
MERDIOINQIVPSMGRTLYTEANQPEVNTAAASDAEGLIENAFQIALNEGSLQV
DKDGVYRIDRYLQIESSDDEYDSPAEEVYITAAKAPQCDIIRHMOGRVIR
YDMAVPLNLSERSAROTGTVGVVLTLSHNSRRIDVDINDONADHRLVIVP
FNDSDVADTQFGSLTRPVNDSAMNNMOGKEAPVPMNMLNYALDGRNGMAVF
SEGLREFEVIGEEKKTPALTLRGVGLIGKEDLLRPGRSCKMPEVPSQRLGSLC
RLSLSTYGTPTAGVACQARALTYOCCNKLPWVMKLNAGEVNPSSYSLMKMP
VGCILSALKKAEDROEYLRLENPAESATCDATVAREVISETMDHEHITTEENO
GSNLSGPFLEGSRTFSYRLA"
4648 . 6216
/gene="cyd1"
/note="cyd1"
4648 . 6216
/gene="cyd1"
/note="ORF_ID:0178#2
similar to SwissProt Accession Number P11026"
/codon_start=1
/transl_table=11
/product="Cytochrome d ubiquinol oxidase subunit I (EC
1.10.3.-)."
/protein_id="BA35399.1"
/db_xref="GI:1651327"
/translation="MDIIVELRLQFALTMAMHFLPEVLTGMAFLIAMEYVVISG
AFPLESTFYGLFEFGMDRLGVOHMCVTMLVAGSNLSALMLIVANGMPLADELMA
KOIKDMTKFWGKLFINFALVATGILTEPEFGFQFNVSYHYGDITGACVIAEGLMA
FETMRKENVSESELYINPYAQVRYHTVASGVTGAMFLIGISAMYMLKGFDPFARK
SEFIAASRGMAYLSTYVLGDSGYEMGVOKTLAAJEAEMEQPAPAAATLGIIDP
OBEETNKFAIQIPYALGIIATRSVDTPVIGLKEIMVQHEERIRNOMKAYSLEDRSG
STDQAVRDQFNSMKRDLGIGLLKRYTPNVADATPAQIQOATKDSIPVAPRYAFRT
MVACGLILAIILAFWSYIRNRIGEEKMLRAALYXGIDPLPWIAVEAGHVPAAETGRP
MAIGEVLPRAVANSILTAGDLIFSMVLICGLTYTLFVLELTMKFAELGPSLSLTKGR
VHREOSTTTOPAR"
6232 . 7371
/gene="cyd2"
/note="cyd2"
6232 . 7371
/gene="cyd2"
/note="ORF_ID:0178#3
similar to SwissProt Accession Number P11027"
/codon_start=1
/transl_table=11
/product="Cytochrome d ubiquinol oxidase subunit II (EC
1.10.3.-)."
/protein_id="BA35400.1"
/db_xref="GI:1651328"
/translation="MIDEVLRFTIMWLIVGVLLIGFAVTDGFDMGVGLMTRFLGRNDT
BRKIMINSIAPHMDGNQVMLITAGGALFAAMPYAAAFSGFYAMLVIASIFPPYV
GEDYRSKILETRRMNMMDWGLIFGSPPLVIGAFGNLLOGVPENDELTARLYYGN
FEOLNPLGGLAGVYVSGMLITOGATYLOMRTVGEALTLRRATQAVALTYTIGFLA
GVWVYGDIGYVKSSTMDHYAASPLMKKEVYREGAGLVFNNTPIIMALPAGVLP
LITLITARMKAAAFVPSSTLACITLTPGAIAPFVPMVSTMMNASLTMMDATSSO
LTINMTAVAVVLPPIILLITAMCYMKMFGRIKEDIERTHTSY"
7499 . 7792
gene

CDS
/gene="ybgp"
7499 . 7792
/gene="ybgp"
/note="ORF_ID:0178#4
similar to SwissProt Accession Number P37343"
/codon_start=1
/transl_table=11
/product="Hypothetical protein in cydB-toIq intergenic
region precursor."
/protein_id="BA35401.1"
/db_xref="GI:4062320"
/translation="MSKTIATLYAVMDKRPRLASPMVLLIAGCMFMPSPRAKTS
ELEIHWGLLMMVAVCAVHVGFRPQKVLMOGIFCPDLADIVLYGLIFFE"
7942 . 8346
/gene="ybgc"
7942 . 8346
/gene="ybgc"
/note="ybgc"
/note="ORF_ID:0178#5
similar to PIR Accession Number A25980"
/codon_start=1
/transl_table=11
/product="15.5k protein (tolA operon 5' region)."
/protein_id="BA35402.1"
/db_xref="GI:4062321"
/translation="NMNTLEFMRPVRYVEDTDAGVYVYHASYVAREARKEMLRHH
FSOQALMAERVAFVRKMYEYVAPARLDDMLEIOTETISMGTSLVFTORIVAEHT
LNDPAVLYVCVPLMKPRALPKSTVAEFG"
8343 . 9035
/gene="fii"
/note="fii"
8343 . 9035
/note="fii"
/note="ORF_ID:0178#6
similar to PIR Accession Number B25980"
/codon_start=1
gene

Alignment Scores:

Pred. No.: 1,77e-19 Length: 16419
Score: 285.50 Matches: 67
Percent Similarity: 56.90% Conservative: 32
Best Local Similarity: 38.51% Mismatches: 72
Query Match: 33.16% Indels: 3
Gaps: 2

US-09-674-779b-2 (1-172) x D90713 (1-16419)

QY	1	MetMetLeuHisIleGlnIleAlaAlaAlaAlaLeuSerValLeuThrPheMet	20
DB	12257	ATGCAACTGACAAAGTCTGTAAGGCGTGATGATGCTGCGCTGATGCAATTGCG	12316
QY	21	ThrGlyCysAlaAspLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly	40
DB	12317	GCATGTTCTTCCACAGACAGACCCAGCAATGACGCGAAGGCATGCTGGTGGCCGC	12376
QY	41	TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrVal	60
DB	12377	---ACTGTATGATGAGCGAAGCGCGCAACAGTCTTCCGAAAGCAGCTCGT	12433
QY	61	AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu	78
DB	12434	CTGCAATGCAACAGCTGACAGCAGCAACATCGTTACTGATCGTGGACAACTAGAT	12493
QY	79	IleLysProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGln	98
DB	12494	ATCCGTTCTGACTTGGCTCAATGCTGGATGGATGCACATCTCCGCTAGCAACCG	12553
QY	99	ThrAlaArgValIleuValAlaGlyHisThrAspGluArgLysSerArgGluTyrTrpAsnMet	118
DB	12554	TCTTCAAAAGCACCGTGAAGGTCACGCGGACCAAGTGTACTCCGGAATCAACATC	12613
QY	119	SerLeuGlyGluArgAlaValAlaValAlaArgAsnTyrLeuLeuGlyLysGlyIleAsn	138
DB	12614	TCCCTGGGTGAAGCTGCGCAACGCGCTTAAGATGATCTGCAAGGTAAAGCGCTTCT	12673

Oy	139	GlAlaSerValGluIleLeuSerPheGlyGluGlnArgProIleAlaPheGlyThrAsn	158
Db	12674	GCAAGACCAGATCCATCGCTTCTTACGGTAAAGAAAAACCGCAGTACTGGGTCAACGAC	12733
Oy	159	GlUGluAlaIrrpSerGlnAsnArgAlaGluLeuSerTyr	172
Db	12734	GAAGCGCATACATCAAAAAACCGTCGCGTACTGGTTTAC	12775
RESULT 19			
AP002553	297816 bp	DNA	linear BCT 07-MAR-2001
LOCUS	Escherichia coli O157:H7 DNA, complete genome, section 4/20.		
DEFINITION	AP002553 Ba000007		
ACCESSION	AP002553.1	GI:13360211	
VERSION			
KEYWORDS			
SOURCE	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)		
ORGANISM	DNA.		
REFERENCE	Escherichia coli O157:H7 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
AUTHORS	1 (sites)		
	Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C., and Shinagawa, H.		
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak		
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)		
MEDLINE	20198780		
REFERENCE	2 (sites)		
AUTHORS	Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.		
TITLE	Comparative analysis of the whole set of RNA operators between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655		
JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)		
MEDLINE	20557356		
REFERENCE	3 (sites)		
AUTHORS	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.		
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak		
JOURNAL	Gene 258 (1-2), 127-139 (2000)		
MEDLINE	20564182		
REFERENCE	4 (sites)		
AUTHORS	Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.		
TITLE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12		
JOURNAL	DNA Res. 8 (1), 11-22 (2001)		
MEDLINE	21156231		
REFERENCE	5 (bases 1 to 297816)		
AUTHORS	Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kengen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)		
COMMENT	genome project.		
FEATURES	Location/Qualifiers		
SOURCE	1..297816		
	/organism="Escherichia coli O157:H7"		

```

/produce="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB3179.1"
/db_xref="GI:13360215"
/translation="MEAKISTEHWYQROKIKKEVTRVAALIEKKKGLITVFETGKQ
KSTAFEGVTVAVGHGKTVAGVQATIKQOMNGEYNLLQPGVFHNGISFTMETGN
QADIDAAEVSSEKSRMLADKRYLVLDLDTVLMIAHYDITDEVTASLQNRPAQS
IATGKGESQILLAKMADYVSELRPNVKNAFDNGIQAPQIDW"
/complement(4332..5984)
/gene="ECS0757"
/complement(4332..5984)
/gene="ECS0757"
/complement(4332..5984)
/gene="ECS0757"
/feature="note=probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
gi121203981|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"

```

```

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFQDQEPFLONKGGTEYTLISDHTITVELDGEVYITAFPE
ALTLTAQAYEAFSEFLRSHILDOVALINDQASANDKYVALQILNAAEVSAGVLTLE
NCDGTGATVIAKSGOQIITGGGADLGGGKGLKSTVPEENLLEFSSQALFLITVNTLT
TNIPADIDISAVAGDEYHETLVKNGGGSKANKALVQETKSLQEPKLAFLITKMSKIS
GTACPFYTHIAEYVGGSLIDYTLKTLAKTASTGYNLLPTSGNGOQAFRDLTEKVLIL
EASDQEPYHIAEYVGGSLIDYTLKTLAKTASTGYNLLPTSGNGOQAFRDLTEKVLIL
EKLEENHGQYTPSLSEENNAHVOLDLRLPRLDVMOLARLPYCTRKSLGPIYVAKR
DIDAHAKIKALDSEENNAHVOLDLRLPRLDVMOLARLPYCTRKSLGPIYVAKR
QAAAGSYLMSKSKNSRSDQYTDACKHKGKGNLSDISGGAALLQAEYVSLRCLEYPDELGA
MEVWMMVYENLPAFLIIVDYKGNFSPQEPQDHRASCAPGH"
/complement(6092..7372)

```

```
complement(6092. .7372)
/gene="ECS0758"
complement(6092. .7372)
/gene="ECS0758"
```

```

/Note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
gi11214671/sp1P24943|GLUT_BACST percent identity 38 in 416
aaa, also similar to C4-dicarboxylate transpor".
(codon_start=1.

```

```

/evidence=not_experimental
/transl_table=11
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"

```

TTEFLRIRKINIPYVUSHLVNGVAMGKGAKIGRIVSNTLPAFCASJLSTALCLOIT
 TIFRIGCSTYALCEGSSALYHLSHSDHAFKLVYSAFPTSDVDMANHEIIOVSEI
 TIGVAGSAGIKPEKPYSTLASSVLPIGSLTVYVGRIRETNATSEAPALFTTSSSAAFP
 TIGTLTIRLBOGNSPKLSSVLPIGSEFNLSGMSGASFPATVYIADACHHISIGIOTI
 TILMLTISGKAMGKGRVARSNVYIATLNLNIPEGLLILMGVDFPLDMGRATVMWNSN
 TIRGALAKPENSANLP

```
complement(7533, .7850)
```

```
gene="FC50/59"  
note="unknown"  
/codon_start=1  
evidence=not_experimental  
tranyl_table=11  
product="hypothetical protein"  
protein_id="BAB34182.1"
```

Alignment Scores:	
Pred. No.:	7.29e-18
Score:	285.50
Percent Similarity:	56.90%
Best Local Similarity:	38.51%
Query Match:	33.16%
DB:	1
	Gaps: 2
	Length: 297818
	Matches: 67
	Conservative: 32
	Mismatches: 72
	Indels: 3

US-09-674-779B-2 (1-172) x AP002553 (1-297816)

24093 ATGCACCTGCACCAAGGCTGATGATTGCTCTGCGCTTATGGCAATTGCG 24152

[illegible]

QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60
|||||
TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60

Db 24213 ---ACTGATATGATGCGAAGCGGCGCACCGCACAACATGTCTTCCGAAGAGACAGGCTCGT 24269

61AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu 78

[illegible]

Db 24330 ATCCGTTCTGACTTCGCTCAATGCTGGATGCACATGCAAACTTCTCGCGTAGCAACCG 24389

[illegible]

OY 119 SerLeuGIygluaRgArGalavalAlaValaIArganTTrleuLeuGIyLysGIIleasn 138
|||||.....
.....GAGCAGACGGCGAAGCGAACGIGeIACITCCGGAAATACAACACATTC 24449

Db
24450 TCCCTGGGTGAACGTCGTGGCAAGCCCTTAAGATGTACTCTGCAAGGTTAAAGCGCTTTCT 24509

[illegible]

159 GIUGLALATrpSerGlnAsnArgarglaGluleuSerTyr 172
| | | | |
.....C..C...CCACAGAACGGTCATGCAC 24569

24570 GAAGCGCATACTCCAAAACCGTCGTGGGTTACTGGTTTAC 24611

AE013906/c		DNA	1 kb	25	700	5000
LOCUS	11138 bp					
AE013906						
RESULT 20						
AE013906/c						

DEFINITION	Yesinia pestis KIM section 306 of 415 of the complete genome.
ACCESSION	AE013906 AE009952
VERSION	AE013906.1 CT-2105006.1
DATE	26-JUL-2002

KEYWORDS
Yersinia pestis KIM.
SOURCE
DOI:10.1007/s12076-011-9150-1

ORGANISM
Yersinia pestis KIM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.

REFERENCE
AUTHORS

1 (bases 1 to 11138)
Deng, W., Burland, V., Plunkett, G., III, Boutin, A., Mayhew, G.F.,
Liss, P., Datta, N., Neidhart, J., and Neidhart, J.

Barry, J., Fehda, N. I., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.

TITLE Blattner, F.R. and Perry, R.D.
J. Bacteriol. 184 (16), 4607-4611 (2002)
JOURNAL Genome Sequence of *Yersinia pestis* KIM

PUMED 12142430
 2 (bases 1 to 11138)
 Deng W, Burland V, Plunkett G, et al. (2002)

Barclay, V., Fairhead, G. III, Boutin, A., Mayhew, G.F., Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., Fetherston, J.D., Linder, L.E., Brubaker, R.R., Plana, G.V., 1998.

TITLE	Author
Straley, S.C., McDonough, K.A., Blatner, F.R. and Perry, R.D.	Nilles, M.L., Matson, J.S.,
Direct Submission	

Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
Henry Mall, Madison, WI 53706, USA
Location/Qualif: f:orc

```
source
location/Vaccinella
1. .11138
/organism="Yersinia pestis KIM"
```

```

/strain="K1M"
/db_xref="taxon:187410"

```

```

gene
    177..512
    /gene="psif"
    /note="y3047"
    177..512
CDS
    /gene="psif"
    /function="phenotype; central intermediary metabolism:
    phosphorus compounds"
    /note="residues 1 to 111 of 111 are 54.46 pct identical to
    residues 7 to 112 of 112 from E. coli K12 : B0384;
    residues 1 to 111 of 111 are 53.57 pct identical to
    residues 1 to 106 of 106 from Genpept : >gb|AA19338.1|
    (AE008713) induced by phosphate starvation [Salmonella
    typhimurium LT2]"
    /codon_start=1
    /transl_table=11
    /product="putative starvation-inducible protein"
    /protein_id="AA06597.1"
    /db_xref="GI:21959962"
    /translation="MRLPMLLSAGLLSANNWADPPKTPSPAAQAOQMTDQNO
    OASTSLKDERKNMSOCLKAQAPDGKALTPQOQKMSQNAQAQKMLGDERKTF
    MSTCLKRAA"
    775..1611
    /gene="gpmA"
    /note="y3048"
    775..1611
CDS
    /gene="gpmA"
    /function="enzyme; energy metabolism, carbon: glycolysis"
    /note="residues 29 to 278 of 278 are 87.99 pct identical
    to residues 1 to 250 of 250 from E. coli K12 : B0755;
    residues 29 to 278 are 89.99 pct identical to
    residues 1 to 250 of 250 from Genpept : >emb|CAD05220.1|
    (AL627268) phosphoglycerate mutase 1 [Salmonella enterica
    subsp. enterica serovar typhi]"
    /codon_start=1
    /transl_table=11
    /product="phosphoglyceromutase 1"
    /protein_id="AA06598.1"
    /db_xref="GI:21959963"
    /translation="MRLTERTORTCSLKYGSHIPLNIGVWMAVTKLVLRGESOM
    NNNRTGTYVDVLSKRSKKAAGKLLKDEGFPDPAPATSVLKRAIHTLMTNLDEL
    DQAMLPTEKTMKLNKHYALOGLNKSEAEKYGDEQYKORRGPAITPLPRLSDSR
    FPGHDPYRAKLTDAELPTTESLATIERTVPMNVNIPRIASGERVIAAHGSLIA
    LVKYLDELGEDELLEINPTGVPLVYEEDENKPKIKHYVIGNADETIAKAAVAVNOGK
    AK"
    complement(1793..2845)
gene
    /gene="argG"
    /note="y3049"
    complement(1793..2845)
CDS
    /gene="argG"
    /function="enzyme; amino acid biosynthesis: Phenylalanine"
    /note="DAHPh synthetase, phenylalanine repressible;
    residues 1 to 349 of 350 are 86.24 pct identical to
    residues 1 to 349 of 350 from E. coli K12 : B0754;
    residues 1 to 349 of 350 are 87.10 pct identical to
    residues 1 to 349 of 350 from Genpept : >emb|CAD05217.1|
    (AL627268) phospho-2-dehydro-3-deoxyheptonate aldolase
    (DAHPh synthetase) phenylalanine repressible [Salmonella
    enterica subsp. enterica serovar typhi]"
    /codon_start=1
    /transl_table=11
    /product="3-deoxy-D-arabinohexulosonate-7-phosphate
    synthase"
    /protein_id="AA06599.1"
    /db_xref="GI:21959964"
    /translation="MNYONDRLRIKEIKELPPVALLERFPSTVNAETVAKTRNAIH
    NILAQRDRLIVYPCSHIDTOAKAYARLALREELQGELEVVRVYFEPKPTTV
    GWKGLINDPHMNSYDINEGRLARLRLIVEINSGLPAGEFLDMITPOLADMSWG
    AIGARTESOVHRELASGLSPVGFKGTGTGTLKVAIDALNASSAPHCSTYKMGHS
    AIVTAGNDCCHILIRGKEPNSSKRVNVAKSLIKAGLEPQIMIDFHSANSCFO
    KOMEVGTVCQIAQGEKSIIGVIESHVEGSONLESGLTYGOSTIDACIGMKDT
    EALLRLAGAVARRRQ"
    3294..4232
    /gene="y3050"
gene
    3294..4232
    /gene="y3050"
    /function="putative transport"
    /note="residues 4 to 310 of 312 are 66.12 pct identical to
    residues 9 to 313 of 313 from E. coli K12 : B0752;
    residues 9 to 309 of 312 are 67.44 pct identical to
    residues 14 to 312 of 312 from Genpept : >gb|AA19697.1|
    (AE008731) putative CDF family transport protein
    [Salmonella typhimurium LT2]"
    /codon_start=1
    /transl_table=11
    /product="putative transport system permease protein"
    /protein_id="AA06600.1"
    /db_xref="GI:21959965"
    /translation="MAVSTIFSDQSNKRLIIAFAITTLFVNTAIGWLSGSLALLA
    DAGHMLTDSALFLIAMAHFESQRPDPRTFGYRLTTLAFAVNAALLIVLITW
    EAHRTFSPHEWKGTPMLIITAGILANIECEWILHKGEEKINVRAAALHVSIDL
    GSVGAMTAIAIVILITGMTPTDPLISLVISVILIRSMRLKESFHELEGAPQEDIN
    KLKRDICTNITVEKNIHIVLMQVGGDRMLTLHAQVTPPLDALLQRIQDYLHMYR
    ISHATVOMEYOHCGTDPDGTINQAPADGHHRRHHE"
    complement(4347..5072)
gene
    /gene="pnuC"
    /note="y3051"
    complement(4347..5072)
CDS
    /gene="pnuC"
    /function="transport; transport of small molecules;
    Nucleosides, purines, pyrimidines"
    /note="residues 1 to 238 of 241 are 76.05 pct identical to
    residues 1 to 238 of 239 from E. coli K12 : B0751;
    residues 1 to 241 of 241 are 100.00 pct identical to
    residues 1 to 241 of 241 from Genpept : >emb|CAC89971.1|
    (AJ414146) integral membrane NMN transport protein pnuC
    [Yersinia pestis]"
    /codon_start=1
    /transl_table=11
    /product="nucleoside/purine/pyrimidine transporter"
    /protein_id="AA06601.1"
    /db_xref="GI:21959966"
    /translation="MDRLSNGNIIIVHPIAGAGYDLSMIEAIGTLFGLCTWPAKRE
    IINYLEGLINVLAVITFQOLIVASLLIQEPFGANIVGVAMSKOTPNQAEKIR
    WLSLPRKALAMAVIAGIIVLMTLHIDIVFAITLVANTVWQSGYANQOMPELDPEP
    FWDSTMYVLSIAAMITMTKRYEWNMLIWNVIDVLSVAIFATQGYAMALEYATLTLIA
    LNSGTMWIKISARNSRPLSTOG"
    complement(5194..6255)
gene
    /gene="nada"
    /note="y3052"
    complement(5194..6255)
CDS
    /gene="nada"
    /function="enzyme; biosynthesis of cofactors, carriers:
    pyridine nucleotide"
    /note="residues 1 to 346 of 353 are 82.08 pct identical to
    residues 1 to 346 of 347 from E. coli K12 : B0750;
    residues 1 to 346 of 353 are 82.94 pct identical to
    residues 1 to 346 of 347 from Genpept : >gb|AA19695.1|
    (AE008731) quinolinate synthetase, A protein [Salmonella
    typhimurium LT2]"
    /codon_start=1
    /transl_table=11
    /product="quinolinate synthetase"
    /protein_id="AA06602.1"
    /db_xref="GI:21959967"
    /translation="KSEITFDVNAAIYPPFAPVPIDTNEKAFYREKIKTLKORDAVL
    VAHYTPDEIOALAEETGGCVADSELMARFGNNHASTILVAGVFMETPKIINPEK
    KVIMPTLNECSIDLCCPVDEFATCDSDPDPTVVAVYATSAVAKADWVYTSIAY
    ELIEHDSIGERITMAPDRHLISGYVSKGAGAVLWQGCIVHDEFKIQALARKKALI
    PDAEIVYHPSPQAVVMDAVGSTSOLIOAKTLPOKTLIVATDRGILFYKQOACPD
    KEIPEAPTAGECATCSGACHPMANNGRAIAEGLEGGVVHEIHDVEILOQALIP
    LNRMLDANOLKLVQKGN"
    complement(6622..6691)
    /product="trna: OTHER"
    /note="anticodon: TTT"
    /function="trna: trna"
    complement(6725..6797)
trna
    trna
    trna

```

/product="tRNA-Lys"
 /note="anticodon: TTT"
 /function="tRNA: tRNA"
 complement(6828..6900)
 /product="tRNA-Lys"
 /note="anticodon: TTT"
 /function="tRNA: tRNA"
 complement(7091..7900)
 /gene="y3053"
 complement(7091..7900)
 /note="residues 1 to 269 of 269 are 69.62 pct identical to residues 1 to 263 of 269 from E. coli K12 : B0742;
 residues 1 to 269 of 269 are 76.64 pct identical to residues 1 to 274 of 274 from Genpept : >emb|CAC82711.1| (A1297885) YbgF protein [Pectobacterium chrysanthemi]"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAM86603.1"
 /db_xref="GI:21959968"
 /translation="MNSNFRHLVGLSLVGVNPMATQAPISNVSGSVEDRYTQ

Alignment Scores:
 Pred. No.: 1.99e-19 Length: 11138
 Score: 283.00 Matches: 66
 Percent Similarity: 56.32% Conservative: 32
 Best Local Similarity: 37.93% Mismatches: 68
 Query Match: 32.87% Indels: 8
 Gaps: 2

US-09-674-779b-2 (1-172) x AE013906 (1-1138)
 QY 1 MethetleuHisIleGlnIleAlaIleAlaIleAlaIleAlaIleuSerValIleuThrPheMet 20
 Db 8416 ATGCACTGCAACAAAGCTAAAGGCTATGTCCTTGGCTTCTGCTGCGG 8357
 QY 21 ThrGlycylAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
 Db 8356 GCTTGTGATTCACAAAGCGCAAAATATGACCAATCTGGCATGGCGCTGCGCTGCTGCT 8297
 QY 41 TTTThrGlyValIleThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60
 Db 8296 -----ACAGAGACGGAGACACCTGCTTCTTAAGGAGCAAGCAGCT 8255
 QY 61 AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu 78
 Db 8254 CTTGACATGCACAAAGCTGCAAAAGAACACATGCTTACTTGGCTTGCATATAATACGAT 8195
 QY 79 IleLysProGlnAlaIleAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGln 98
 Db 8194 ATGCGTTCGACTTCGCTCAAAATGCTGATGCACATGCTGCTGCTGCTGCTGCTGCTGCT 8135
 QY 99 ThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlyLysThrAsnMet 118
 Db 8134 TCTGACAAAGTGTGTTGAGAGCCGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8075
 QY 119 SerLeuGluValArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsn 138
 Db 8074 GCTTGGTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8015
 QY 139 GlnAlaSerValGlnIleLeuSerPheGlyGluGluArgProIleAlaPheGlyThrAsn 158
 Db 8014 GCTGACACAGATCTCATGCTTCTTACGCTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7955
 QY 159 GluGluAlaIlePheSerGlnAsnArgAlaGluLeuSerTyr 172
 Db 7954 GAAGCAGCTTGTCTTAAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7913

RESULT 21
 AJ414146 210050 bp DNA linear BCT 06-JUN-2002
 LOCUS
 DEFINITION Yersinia pestis strain CO92 complete genome; segment 6/70.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Yersinia pestis.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Yersinia.
 1 (bases 1 to 210050)
 Parkhill,J., Wren,B.W., Thomson,N.R., Titchell,R.W., Holden,M.T.G.,
 Prentice,M.B., Sebaihia,M., James,K.D., Churcher,C., Mungall,K.L.,
 Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdano-Tarraga,A.M.,
 Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,
 Fairley,S., Hamlin,N., Holroyd,S., Jagsels,K., Leather,S.,
 Kariyasev,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K.,
 Simmonds,M., Skellon,J., Stevens,K., Whitehead,S. and Barrrell,B.G.
 Genome sequence of Yersinia pestis, the causative agent of plague
 Nature 413 (6855), 523-527 (2001)
 21470413
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 source
 1.210050
 /organism="Yersinia pestis"
 /strain="CO92"
 /db_xref="taxon:632"
 /note="biovar: Orientalis"
 829..11946
 /gene="yp01004"
 /note="yapH"
 829..11946
 /gene="yp01004"
 /note="yapH"
 /note="Previously sequenced as Yersinia pestis YapH
 protein TR:CA14227 (EMBL:A127631) (3705 aa) fasta
 scores: E(): 0.99.6% id in 3705 aa. Similar in regions to
 Escherichia coli ycha protein TR:09JMS3 (EMBL:AP001918)
 (1371 aa) fasta scores: E(): 0.39.1% id in 1405 aa, and
 to Escherichia coli adhesin Alda-1 precursor SW:AI04.ECOLI
 (Q03155) (1286 aa) fasta scores: E(): 0.29.2% id in 1286
 aa. Contains a possible N-terminal signal sequence."
 /codon_start=1
 /transl_table=11
 /product="putative autotransporter protein"
 /protein_id="CAC69847.1"
 /db_xref="GI:15979073"
 /db_xref="SPTREMBL:O8ZHA1"
 /translation="MTIFKVIYNASLVNVAIVYSELAKGRIKRSRNLISEGVLPK
 EOSWSKLFKRNLIATLSGLIVELSTGSPADIVYTSQKLEALASNLISGVLPK
 DITLIGLIVNMSNOVVIDGCGFGLTVNNTNKGIVSSGSGTFLPQMSKIDSDAN
 YTSVNAFAVTFADGAFELIANQWNSGTTNRGLETGSGNSIDFGSGVKIVLSR
 SDGSIISGNDIDNATNAGIINWASDANVINYLGSLIKATNIGLATKANNAS
 DIVIRAGDITTAATGISTANGGTGIVKNDGTTSTGATLISASIREISDNDG
 TITATAGTGVNVALASATLNEFGTINTSATANGITPAGCGGTITDNLGNTGTA
 LSNVAGNLTLSNVTTLNLTNGTALNSLTGLTIVDSUNGRTINIEAGIGIATNTEL
 NTPAEALDINVAGAGIQTGCGVLSNLIINVAANTLGTALDITDGIIDNTTIG
 NEIOLANENATINFLASSSKTLNNNGTIRKGSVIFAGVADHIINNGTLDGLITGAG
 NDITLVDSSQSQSDVNLGDGNSVTIOMGATVSSITLNGNDTFTINMGSVSTYLG
 SLDAGTGLNTLNASTDELAATSLQGFNLINLVDSTLIDNDGIDGSGNIDSS
 ELFGFSTFDILHATLGAGTGSALVNNASVLEOAMFAGTMOVNOGALPANSNO
 LGSAGIGIDGTLNDNLALFNHVLNGNTLVNKAATAPAGDSYVGAFSGIYNLT
 KTYTALSDANAAASATLTKSDSYTYGTTDRLTGHGIDLPGAGVAPQSQTS
 GYTVTIDIALINSGTVNITSGSDMDNDPLATNVSILEQDASSTLELINATVYTDID
 ALDLVNGRIITISGTGVSALIQGGSTVAAIHNYGLASSNSNGSGLYVNTLSAL

TEFLMKLGRSDYEGFDPKSSKAGRAMGOISGMPVLLLEADSEDPDKAKTFFEND
ELKDFEGGTLATRGVNGNDYEPFPGTIVISONAAVDASEAILIRIVKLPKRP
OVTESRISADNLNLOVEELSHLIKAVCEGATLEKFAERKFEYERLEKEDRL
EVIYKNNHOMALLDICIRVITIPPEMKATRDALLEAFEROKAISADHVOVEMFE
VVEYLATGNGKPVNHSRDSRIAININORAKAKAOSQVYPDLKIVIRGLAASRRH
KLVSANTAVNSAVLTGFGAGTTVCWFSK
4103. .4327
/gene="XCC3003"
4103. .4327
/note="putative; ORF located using Glimmer/Genemark"
/gene="XCC3003"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA042275.1"
/db_xref="GI:21114215"
/translation="MMOGIVLMLHSDQHDVAPVLINCEATGFDVLLGGDAHSVL
GRHVCMRVLDVLTWLOKORANGAR"
4324. .4587
/gene="XCC3004"
4324. .4587
/note="putative; ORF located using Glimmer/Genemark"
/gene="XCC3004"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA042276.1"
/db_xref="GI:21114216"
/translation="MNSVRATVPEADDTISREHDIRMVOQAASLILNHDIA
RAGISHDGLAAVADFMRELLDIACNAHREPTKPTGADLI"
4666. .5076
/gene="XCC3005"
4666. .5076
/note="putative; ORF located using Glimmer/Genemark"
/gene="XCC3005"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA042277.1"
/db_xref="GI:21114217"
/translation="MOOHAVTRSOISAGGQATTPVEAFDLAVGKDCSALATLY
THDAVVVAALITMGHSSAONRWERRRGPKGKILSGPLFTSEAHRTISMLAKFMD
DLDPEFLANMLPRPTAAEAIAAAREVANA"
5090. .5245
/gene="XCC3006"
5090. .5245
/note="putative; ORF located using Glimmer/Genemark"
/gene="XCC3006"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA042278.1"
/db_xref="GI:21114218"
/translation="MIAPTIGALLYRLWISPTPTAHIGLAVGQIPOLRRRRAMA
VRRVYAHG"
5238. .5513
/gene="XCC3007"
5238. .5513
/note="putative; ORF located using Glimmer/Genemark"
/gene="XCC3007"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA042279.1"
/db_xref="GI:21114219"
/translation="MAISIVYGMASGSLNABICQAVGLKRVLEDRLOKRGD
WOLSQNDVYMLTNDQALAEKTAQRKRTVAITARLRVGAAMALR"
5606. .5755
/gene="XCC3008"
5606. .5755
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1

US-09-674-779b-2 (1-172) x AE012414 (1-12393)

Alignment Scores:

Pred. No.:	1,11e-18	Length:	12393
Score:	276.50	Matches:	63
Percent Similarity:	55.09%	Conservative:	29
Best Local Similarity:	37.72%	Mismatches:	160
Query Match:	32.11%	Indels:	5
DB:	1	Gaps:	5

gene
AE011958
/gene="XCC3009"
5932. .6024
5932. .6024

transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA042280.1"
/db_xref="GI:21114220"
/translation="MLNRHDLMDMDYTALQALDRIDEDTTAHLROAERAINNGAGA
EENYA"
5932. .6024
/gene="XCC3009"
5932. .6024

RESULT 23
AE011958/c 11713 bp DNA linear BCT 29-MAY-2002
LOCUS
DEFINITION
Xanthomonas axonopodis pv. citri str. 306, section 336 of 469 of
the complete genome.
ACCESSION
AE011958 AE008923
VERSION
AE011958.1 GI:21109461
KEYWORDS
Xanthomonas axonopodis pv. citri str. 306.
Xanthomonas axonopodis pv. citri str. 306.
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xanthomonas.
1 (bases 1 to 11713)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,


```

/codon_start=1
/transl_table=11
/product="phosphoserine aminotransferase"
/protein_id="BAB50673.1"
/db_xref="GI:14024069"
/translation="MTTHRRPGLRPNPNFSSGPPAKKRPQWSVEALKNALGSRSHRAK
IGTKRLEOAIETREILQVADHRIGIVPASDVGAEVMAIWSLGERGVDMVAMESEFG
SGWVDVVKQLKLPDVKRLKLEAGYELPDLAKIDEDRDVVTWNGTSGVAVPDEIFP
ADRRGLTICDATSAAFORLDEKLDVTFESMQKVLGGAGHGMILISPRAVARLESY
KPMAPLPKIFRLTSGGKLEIGIKGEFINTPSMLCYEDYLDALHMAKSGTIGDIYAR
ADANAVALDFEYKSSMLGIAVOPATFRSNTSYCLSTPTPDVAALDADQATPAKGLV
SALDKGAVLDIYSYSDSDPGLKINMGCAIYERSDLETLPLDMWAPAOKASILKAA"
complement(3096..3869)
/gene="m113877"
complement(3096..3869)
/gene="m113877"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50674.1"
/db_xref="GI:14024070"
/translation="MTLSRIYVLAALGLPLTPAVADYDPPIYADQAPPVPEV
GSGWYLRGVSYLAOKSFKNDDPAFTPAFSEDEKEDPIEASIGFYHENDYLAADNLG
VGLGNKIGIGYDSDLSVPKDPATSVASADLKNAYSIMLNATVDGTYGITYLGCG
VGVYOSTRRLSANYPFNKGDPTDDEVDYDOKTKYSLAYTLNGLAVQVSKNSVDLY
QTFAPDAEYVNASLTSFPPKHSINHQIKLGRYDL"
complement(4111..4944)
/gene="m113878"
complement(4111..4944)
/gene="m113878"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50675.1"
/db_xref="GI:14024071"
/translation="MNTARKALLALIFAGLASPYPAADLPPEVEAPPYEEDPV
VGGWYLRGIDYHKSVDYRGIDYMTYCIDPCNCTVAPGKSPDYGLKGFSLGSGVY
KINDRLRADYVADIMFKSNFNGSTDLVSTSEVSKMSALLLANAYVDIGTWGITFP
YVAGIGGARVYMDYVDPNTAETNPASNMRFPAALMAGASYCLTDKILIDAGRFS
HIOGGRMFEMDASSAGPFGDNGINTHEVAGLRQYOGNGGCAAVAVAYQPEPEIYR
K"
complement(5193..6545)
/gene="m113879"
complement(5193..6545)
/gene="m113879"
/codon_start=1
/transl_table=11
/product="phosphoglucosyltransferase/phosphomannomutase family
protein: MraA"
/protein_id="BAB50676.1"
/db_xref="GI:14024072"
/translation="MAGNYFGTDGIRGRANKFPMTAETAMRVMAAGLSFQGRSHRH
VVLGDTLRISGMIEINAMVAGIACAGMDVFLGPIPTPAVAMLVRSLLADIGVMSAS
HNPYDNGIKILFPGDYKLSDEIERISMDKDELVELTSDGIGRAKRDYCHDR
IEFAKRTLPBSMSLSGLRIYVDCAGAAKYAPPEALMELGAEVAINIEPGEFINSK
CGSTHPAGIOKKVHEVRADICGIALDDGDYVYIDENCAIVDGDQIMAMIAESWQSG
RLAGGCVSYVMSNLGLERFJGDKRLQIHRKVGGRVYEHMRAHGLVAGGEOGSHY
LSDFSTGDSVLSALQIACIKRQGRPSELSKKEPEYDOLKINRIRLNGKPLDEAPY
KAAIEDARNRLKAGRLVIRPSGTEPLIRVMAEGDDPQVAVVNDIVEIVSETRSA
"
6913..7779
/gene="m113880"
6913..7779
/gene="m113880"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50677.1"
/db_xref="GI:14024073"
/translation="MLHRLATLTGLAVVTLTFLAANALTMACSTRYNAXKADGTL
AGQIWNQFRAQCGTDAATATDTRPATTKAETPPAKATETKPPAKAAAADSTAKGL
TSKESAKYQAAKAGATLNGMKWDFRTECAAGASAAADTTKATTKPATTKAAP
AGGKGLTMAECSAKYQAAKTANTLKQKWNDFRKAEGCAAAADDSVPAPSEANYTKE

```

```

PAKPTTVA PKGTFPSAISPKYASSTPKGMMHTCLDQYALKAMALGCIKWIOKGG
GFSYLCNAKIK"
8079..9537
/gene="m113881"
8079..9537
/gene="m113881"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50678.1"
/db_xref="GI:14024074"
/translation="MPNLKLAIVLVTSVANSVAPASALTMKEGGEYRYTAKEGGTIN
GMDMAEFREKCAPPAESVGDYTKTPQKRTDQAVATVAPASVPTPTTLAAEF
KTEKPAKRMKTCLOGYGNKEAGTINLNRWIOKGGTISLCNRLAES"
complement(8624..10552)
/gene="m113882"
complement(8624..10552)
/gene="m113882"
/codon_start=1
/transl_table=11
/product="metalloprotease (cell division protein); FtsH"
/protein_id="BAB50679.1"
/db_xref="GI:14024075"
/translation="MNPYRMLALMAIIVALLILFNLFPOTRGASDPVYSOFIO
DVAAGRYVKTITAGARITGYTDSSTGEFQYSPDQVLSRLDKNTINAREBADS
NSLGGYLSMILPMLILIGWVTEFMRQKSSGRAMGFSKAKLITABGRVTFDVA
GVDAKEDLEETVEFLNDPOKFORLQKIPRGVLLVGPFGTKTLARVAGRANVPE
FTISGSDPEVMEVGVASRVMDPMQKKNAPCIIFIDEIDAVGRHGAOLGAGNDER
BQTLNOLIVEMDGFESNESIILIAATRPVLDALRPPRQVYVPPDILVREK
ILKAVRNVPPLAPVNDLVKVRAGTFRSGADLMNLVNESALMARKKRLVYMAEED
AKDKIMGAERSSAMTOAEKELTVAHAGHATITLNPESADLRRKRLVYMAEED
VMOLPBGDRYSMSYKYMISRLATMGGRVAAEFREKENTTSGASDIEQATLARM
VTRNGFSDKLCHVAVGNDQEBVFLGHSVARTQNTSEETAQIIDEAVRLLIDEAVSTAK
SILTKKKRGTWALQGLLEYFTTSGDEIKOLINGEKPARLDGDDTPSRGSAVPKSSG
RRKKPEEGGEMALPOBSS"
complement(10675..12039)
/gene="m113884"
complement(10675..12039)
/gene="m113884"
/note="Similar to yaen protein [Bradyrhizobium japonicum],
contains similarity to cell cycle protein Mesd/cytosine
deaminase-related protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50680.1"
/db_xref="GI:14024076"
/translation="MDTEPDLSITRFSHIDFTHGAVAAVSGGSDSTALLILKHNFD
RYSKSAKLAVTIDHGLRQGSAAEAQVAKLCAERGIARHTITWTRKPSGTPAAR
DARYHLIAEARAAGIGLIVYGHADQAEVLMRHADELEADLGRGLAGMAPATL
YMRWEMIVRPLGTGRSALRDLRREHVGAEEDPTINDEAFERRPALAGSAGHM
ENTLRLAGOAVERGOLGASANLIRIASQPSGTGLRIDPALLIADDOAATYALRL
LATRAGAAFTPDQARSALALVRLKAGLCLCTSRIVYDREHAFILRRARGLPPAT
AVDNTLMDGRHHTITLNDMSGALLIAPGAAARLALIDGEPASLIRALAAEPILL
QAFENLGLPQGEERLPLVFARPVVSPFARLPFSDIAPARAVALIGASPLPASPLGG
H5AD"
complement(12098..13210)
/gene="m113886"
complement(12098..13210)
/gene="m113886"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50681.1"
/db_xref="GI:14024077"
/translation="MLIRSVLSGTIAALLISGFTLALASGPGASGEGQSTDSGFSPHL
SLPSPILFGDKKRPDOVQFPAQOSNGAGAGLEQDLQKNGKILEINFOILQOQMRK
QOEDNEFRQQLGGAQGGQPPAPKKSATTTNTNDVAAAPATQAPADAGAGGNOS
AGKGTIEDVIVESPEDGPKVPIPGTAGPAGTGSITLNDGNVLIAGGNTOSTAPTD
AAPATKAPAKAGKSNCTIVAAIPSTNDPELVRNSYQFLLSDDYTAEGGFRDHSIRF

```

Alignment Scores:
Pred. No.:
Score:

4.45e-16 Length: 349116
269.50 Matches: 64

Percent Similarity: 55.97%
 Best Local Similarity: 40.25%
 Query Match: 31.30%
 DB: 1
 Gaps: 5

US-09-674-779b-2 (1-172) x AF003003 (1-349116)

QY 13 AlaleuSerValleuThrheMetThrGlyCysAlaAsnLysSerThrSerClnValMet 32
 DB 13885 GCGCTGGTGGCGATGCTGCCATCGCGCGCTTGGTGGATGAGAGAGCG----- 13838
 QY 33 ValAlaProAsn-----AlaProThrGlyTyrThrGlyValIleThrGlyValAla 50
 DB 13837 -----CCGACACACCGCCCGCCGACCTCGGCTCAACAGCG-----GCCGGTGGCGGA 13793
 QY 51 ---ProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerIleu 69
 DB 13792 ACCCGCGGCGCGCGAGCGAGCTTCAACCGTC-----AACATTCGCGCGACCGC 13748
 QY 70 ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGlu 89
 DB 13747 ATCTTCGATACGACGCTGCTCTGCTGATCGCGCGCGCGACCGACCGACCTCGCTCC 13688
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
 DB 13687 CAGCGCGAGTGGCTGAACCGATCAAGACATGACGATCGCTCGTCAAGATCAAGCGCGAC 13628
 QY 110 GluArgGlySerArgGlyTyrAsnMetSerLeuGlyGluArgArgValAlaAlaValArg 129
 DB 13627 GAGCGCGCGACCGCGATCAATCAATCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 13568
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleLeuSerPheGlyGlu 149
 DB 13567 GACTTCCTGCTTCCAAAGTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13508
 QY 150 GluArgProIleAlaPheGlyThrAsnGluGluAlaIleTrpSerGlnAsnArgAla 168
 DB 13507 GAACGTCGCGTGGCGGTGCTGCGAGACATCTCTGCTGCTGCGCGCGCGCGCGCGCG 13451

RESULT 25
 AEO04009/c 10833 bp DNA linear BCT 15-JUN-2001
 LOCUS Xylella fastidiosa 9a5c, section 155 of 229 of the complete genome.
 DEFINITION AEO04009 AEO03849
 ACCESSION AEO04009.1 GI:9106980
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Xylella fastidiosa 9a5c.
 Xylella fastidiosa 9a5c.
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xylella.

REFERENCE
 AUTHORS
 1 (bases 1 to 10833)
 Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
 Alvaranga,R., Alves,L.M., Araya,J.E., Bala,G.S., Baptista,C.S.,
 Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
 Bueno,M.R., Camargo,A.A., Camargo,L.E., Carriro,D.M., Carriro,H.,
 Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
 Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,
 Facinani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
 Fraga,J.S., Franca,S.C., Franco,M.C., Fromme,M., Furlan,L.R.,
 Ganiar,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
 Ho,P.L., Hohnselt,D.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.,
 and Marino,C.L.
 The genome sequence of the plant pathogen Xylella fastidiosa. The
 Xylella fastidiosa Consortium of the Organization for Nucleotide
 Sequencing and Analysis
 Nature 406 (6792), 151-157 (2000)

JOURNAL
 MEDLINE
 PUBMED
 10910347
 2 (bases 1 to 10833)
 Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
 Alvaranga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S.,
 Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
 Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carriro,D.M.,

TITLE
 JOURNAL
 FEATURES
 source
 location/Qualifiers
 1..10833
 /organism="Xylella fastidiosa 9a5c"
 /db_xref="taxon:160492"
 /clone="9a5c"
 192..2081
 /gene="Xf1888"
 192..2081
 /note="similar to SP|P45740 (percent identity: 67 %/query alignment coverage: 91.4 %/subject alignment coverage: 97.5 %) identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="chiamine biosynthesis protein"
 /protein_id="AAF84694.1"
 /db_xref="GI:9106981"
 /translation="MDRAAPRMQRKRRPVPSPISGRTVITATLPIPLILSE
 KVTAPLAKRIYITGSKODIRPMPREIALIPSSARRGNGENSLAXIDSGITDPO
 ATIDLACGLPRLPAMVIDERADYVAALYFVPSPVATPPTPARPRARNAV
 TOLEYARGLVTPPEMEVAIREQORREQVENLNGORAGADGALVGTPTPEFND
 EIRGRALIPNNIHPESEPMIGNPLTKINAMIGTSLSSIAEVEKLVAMIRNG
 ADTINDISTGNDIHATREMIIRNSPVPGVPIVQALEKVGHEALSAWAFRTDLE
 QAEQGVYVYHAGVLRPFITPLPSRLTGIYSGSGSIMARCOAHRSENPLTFEEL
 CEIMAYAVAFSLGDLRPGCLADANDAGRALELDELTHIMNNOYVMVSGPH
 VPMHLIKANMKQLACGAPFPYLTGPTTIDAPYDHTISATIGAMMGITMALT
 WPKHEGLIPNLDVHDGIIIVKIAHNAADLAKGHPAQAQDADLSAREFERMDGF
 HSTPEKALALHDSLEPKHAHRAFCSCMQPQFCSMKISOEVRDASSNLSGNTN
 TNDAT"

gene
 CDS
 2213..2992
 /gene="Xf1869"
 2213..2992
 /note="similar to SP|Q44532 (percent identity: 59 %/query alignment coverage: 97.3 %/subject alignment coverage: 97.7 %) identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="ferredoxin-NADP reductase"
 /protein_id="AAF84695.1"
 /db_xref="GI:9106982"
 /translation="MSPAFTETVYVHNMWDATSFITTDGSGRFGQVWIGLE
 TPARPLRAYSLASNAWEQLEFISIKONGLSLKRHLKRPDKLITGKRPSTLITL
 HDHGRHLVLTGTGLAPWMSIIKDEYTERFQVLTITGVRISKDLARFKE
 LPQHEHLEGTIRKRLIYPAVAREDFPNRGRILTHIESGAMONTGLPIIDQANDREM

gene
CDS
complement(2986..3558)
/gene="XF1890"
complement(2986..3558)
/note="Similar to SP1032770 (percent identity: 50 %/query alignment coverage: 82.6 %/subject alignment coverage: 100.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="glutathione peroxidase-like protein"
/protein_id="AAF84696.1"
/db_xref="GI:9106983"
/translation="MHTVSHAFASKLSLSDRPMSSITPTPTRLDRGQALADH RCOVLIVNVAASRGCTPOYAGLEMLQRYADAGLIVGFCDDPAGDEDEAKLAE FCTLVYGDPEPAKIKYNGADAHPLMOWLHRRRGLFGMAIKMPTFKLIGNRGP IARYSPKSPQLEHYIYALGEBGGNTH"
3884..5416
/gene="XF1891"
/gene="XF1891"
3884..5416
/note="Similar to SP1P94408 (percent identity: 35 %/query alignment coverage: 96.7 %/subject alignment coverage: 100.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="G1-cripeptide ABC transporter membrane protein"
/protein_id="AAF84697.1"
/db_xref="GI:9106984"
/translation="MSVDTTVKNTVYATPQTLPKFTTFGHPKPLMLMAEFWER FAFYGIKALVLTQTFDGNISGEENASTYGAVALYASAIFFGFTADRIQGO RSLIGAVIMACGLITVPSRMEFEGLATVINGLPEKPNISLGGOLYADGDR DRGTLIFMYGNAGSLISPLTSMAGOVCTPOQYKVFYASGGMFLIMPIV GKRQKIGLEPKDESIFFRTFLIFGALAPLAYALLAKNATTTAAILGLIFAL ALITIVLRNGKLTORDRYATLIIFFENFMFEPQAGSSSENFKNKLVDSQILSN WEPVGMOSNPLAIIILLAPILIVIMSILDKRIEPISTPKSLGMFNGFGLIM YASNLNLSNMTPFMSLIANYIVQVGECLSPITGSMVTKLAPVLVGFAMGWFEL STRAGNLSGVFAFVSVEEGMTVSSALRGYTFGWSLIGSILLFLISPLINRHHG VK"
complement(5578..6342)
/gene="XF1892"
complement(5578..6342)
/gene="XF1892"
/note="Similar to SP1P32679 (percent identity: 53 %/query alignment coverage: 83.1 %/subject alignment coverage: 94.6 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="endonuclease V (deoxyinosine 3'endonuclease)"
/protein_id="AAF84698.1"
/db_xref="GI:9106985"
/translation="MPVYALISDTGMVFMFKISSIDIFAGMDGSTTEARLOSMAE RYVLKDLNLSLEPTLAGDVGFEDEGTRAAAYAMACDKILETVVYPTSPV YVGLSFRELPAALQALQLSRIPALVVDGGAHPRGLGAAHGLVTLNPGCV AKRRIVGDEPCTAGERTPHILHCTGVGMALRSKIRCKPIMISPGHKISLSHALTW TORCLNGRLEPPTROADRLASRGOKIVSDLSL"
6397..7146
/gene="XF1893"
/gene="XF1893"
6397..7146
/note="Similar to SP1P31217 (percent identity: 56 %/query alignment coverage: 98.8 %/subject alignment coverage: 98.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="phosphoglyceromutase"
/protein_id="AAF84699.1"
/db_xref="GI:9106986"
/translation="MTRKLVLLRHGOSQWNSMNRFTGWVDIGLTQGHQDEATMAGHLM

gene
CDS
complement(7245..7928)
/gene="XF1894"
complement(7245..7928)
/note="Similar to GI15514780 (percent identity: 63 %/query alignment coverage: 93.0 %/subject alignment coverage: 98.1 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF84700.1"
/db_xref="GI:9106987"
/translation="ANDQRNNRNLNATRSQKLSIFLSLOGEANSVGMPTVYRLT GCPRLROYCDTATAFHGEWCSDITIVSEVRSYGRHVCVTGEGPLAKRCLLEKL CDAGFEVSLSTGALDIAVDLRVSRVDIKTFGSEAHNRNHPNIALITLPHDQIKFV LCSRADYEMARTCVAEHELERRCMWFSPSKODIAPTVLADMLISDRIGRRQDLHK LHMNDEGR"
Alignment Scores:
Pred. No.: 1,37e-17 Length: 10833
Score: 265.50 Matches: 60
Percent Similarity: 53.89% Conservative: 30
Best Local Similarity: 35.93% Mismatches: 54
Query Match: 30.84% Indels: 23
Gaps: 5
US-09-674-779b-2 (1-172) x AE004009 (1-10833)
QY 20 MetThgLYcysAlaAsnlySerThrSerGlnValMetVal----- 33
Db 9250 CTTGTCGCGTGTCAAAAAAGTTAAGAGCAACCAACGATGCGCTGCAAACTATGGCA 9191
QY 34 -----AlaPro-----AsnAlaProThgLYTThgLYValle 45
Db 9190 CCACACTGTTTACGCCGCCGCCACCACTACGCGGCCCAACCGATTCTTGGG---CTT 9134
QY 46 TyrThgLYValAlaProLeuValAspAsnAspGluThrVallyAlaLeuAlaSerlys 65
Db 9133 TACACA-----CGCGCTACCTAGTATACCGATGCTTGGTGGT----- 9095
QY 66 LeuProSerLeuValTyrThrAspSheAspSerAspGluLeuLysProGlnAlaAla 85
Db 9094 ---CAACGTGTGCTATTTCACCTTGTATTAAGATGATGTAAGAAAGAAATTCCAACG 9038
QY 86 IleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAla 105
Db 9037 GTTCTTGGTGGTCCATGCAAAATATCTTGGCAACGCTCCCTTGCACACATTCATACG 8978
QY 106 GlyHisThrAspGluArgLYSerArgLutryrAsnMetSerLeuGlyGluArgAla 125
Db 8977 GCCAATAGTGAAGACCGCGGTGCGGTGAGATATATAGCTCGGTGGAACGCCGCGGT 8918
QY 126 ValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIlele 145
Db 8917 AACCTGTGTTGTATTCAATTACAGCGAATGTCCTGTCGGGACAGTTGAAAGTTGTT 8858
QY 146 SerPheGlyGluGluArgProIleAlaPheGlyThrAsnGlnGluGluAlaIleAsn 165
Db 8857 AGTTACGCTGAAGACCGTTCGCTTGCATCAATCAACACAGAAATGCTGTGTCACGTAC 8798
QY 166 ArgArgAlaGluLeuSerTyr 172
Db 8797 CGGCGCTTGAATCGTCTAT 8777
RESULT 26
AR089419 462 bp DNA linear PAT 07-SEP-2000
LOCUS AR089419
DEFINITION Sequence 178 from patent US 5994066.

TITLE VACCINES AND DIAGNOSTIC ASSAYS FOR HAEMOPHILUS INFLUENZAE
JOURNAL Patent: WO 9002557-A 3 22-MAR-1990;
FEATURES Location/Qualifiers
source 1. 737
BASE COUNT 224 a 134 c 154 g 225 t
ORIGIN

Alignment Scores:
Pred. No.: 6,32e-19 Length: 737
Score: 264.00 Matches: 51
Percent Similarity: 68.93% Conservative: 20
Best Local Similarity: 49.51% Mismatches: 32
Query Match: 30.66% Indels: 0
DB: 6 Gaps: 0

US-09-674-779B-2 (1-172) x 106718 (1-737)

QY 70 VALTYRPHASPPHASPERSASPGLULLEYSERPROGIALAALALALEUASPGLU 89
|||||
DB 391 GTTATTTCGGTTTGGATTAATATGACATTACTGGTGAATACGTTCAAATCTTAGATCGC 450
|||||
QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
|||
DB 451 CACGCTGCATATTTAATGCAACACACACCTCTTAAGTATTAGAGAGGTACACATGAT 510
|||||
QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
|||||
DB 511 GAACGTGTGACACAGAAATACACATCGCATTCAGCCACACGTCTGCAGATGATTAA 570
|||||
QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
|||||
DB 571 GGTTATTACCTGGTAAAGGTGGTGAATGCTGTAATTAAGGCACAGATCTTAGCGTGA 630
|||||
QY 150 GluArgProIleAlaPheGlyThrAsnGlnAlaIleTyrSerGlnAsnArgAlaGlu 169
|||||
DB 631 GAAAAACCTGCAGTATTGATGATGATGATGATGATGATGATGATGATGATGATG 690
|||||
QY 170 LeuSerTyr 172
|||||
DB 691 TTAGCGTAC 699

RESULT 29
108799 108799 737 bp DNA linear PAT 02-DEC-1994
LOCUS Sequence 2 from Patent WO 8804932.
DEFINITION 108799
ACCESSION 108799.1 GI:588491
VERSION
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 737)
AUTHORS Deich,R.A., Zlotnick,G. and Green,B.
TITLE VACCINES AND DIAGNOSTIC ASSAYS FOR HAEMOPHILUS INFLUENZAE
JOURNAL Patent: WO 8804932-A 2 14-JUL-1988;
FEATURES Location/Qualifiers
source 1. 737
BASE COUNT 225 a 134 c 154 g 224 t
ORIGIN

Alignment Scores:
Pred. No.: 6,32e-19 Length: 737
Score: 264.00 Matches: 51
Percent Similarity: 68.93% Conservative: 20
Best Local Similarity: 49.51% Mismatches: 32
Query Match: 30.66% Indels: 0
DB: 6 Gaps: 0

US-09-674-779B-2 (1-172) x 108799 (1-737)

QY 70 VALTYRPHASPPHASPERSASPGLULLEYSERPROGIALAALALALEUASPGLU 89

DB 391 GTTATTTCGGTTTGGATTAATATGACATTACTGGTGAATACGTTCAAATCTTAGATCGC 450
|||||
QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109
|||||
DB 451 CACGCTGCATATTTAATGCAACACACACCTCTTAAGTATTAGAGAGGTACACATGAT 510
|||||
QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
|||||
DB 511 GAACGTGTGACACAGAAATACACATCGCATTCAGCCACACGTCTGCAGATGATTAA 570
|||||
QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
|||||
DB 571 GGTTATTACCTGGTAAAGGTGGTGAATGCTGTAATTAAGGCACAGATCTTAGCGTGA 630
|||||
QY 150 GluArgProIleAlaPheGlyThrAsnGlnAlaIleTyrSerGlnAsnArgAlaGlu 169
|||||
DB 170 LeuSerTyr 172
|||||
DB 691 TTAGCGTAC 699

RESULT 30
HEAOMP6 867 bp DNA linear BCT 26-APR-1993
LOCUS H. influenzae outer membrane protein p6 gene, complete cds.
DEFINITION M19391
ACCESSION M19391.1 GI:148960
VERSION
KEYWORDS 11poprotein; outer membrane protein.
SOURCE H. influenzae (isolate 1479) DNA, clone pBUD5.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.

REFERENCE 1 (bases 1 to 867)
AUTHORS Nelson,M.B., Apicella,M.A., Murphy,T.F., Vankeulen,H., Spillia,L.D.
TITLE Cloning and sequencing of Haemophilus influenzae outer membrane protein p6
JOURNAL Infect. Immun. 56 (1), 128-134 (1988)
MEDLINE 88085463
PUBMED 3257200

COMMENT Draft entry and printed copy of sequence for [1] kindly provided by M.B. Nelson, 27-May-1988.
FEATURES Location/Qualifiers
source 1. 867

CDS
/organism="Haemophilus influenzae"
/db_xref="taxon:727"
68..529
/note="outer membrane protein p6 precursor"

/codon_start=1
/transl_table=11
/protein_id="AA024994.1"
/db_xref="GI:148961"
/translation="MNKRVSLVAGSVAALACSSNNDAAGAGAGTGGYSVADL
QORNVYFEGEPKDYDTEYVOILDHAAVYINAPAAVIVEGMTDEGTPEYNIAG
ORRADAVKGYLAGKGVDAKIGTYSVGEKPAVIGHDAAVSKNRRAVLAV"
68..124
/note="outer membrane protein p6 signal peptide"

sig_peptide
mat_peptide
BASE COUNT 253 a 168 c 175 g 271 t
ORIGIN Unreported.

Alignment Scores:
Pred. No.: 7,78e-19 Length: 867
Score: 264.00 Matches: 51
Percent Similarity: 68.93% Conservative: 20
Best Local Similarity: 49.51% Mismatches: 32
Query Match: 30.66% Indels: 0
DB: 1 Gaps: 0

US-09-674-779B-2 (1-172) x HEAOMP6 (1-867)

OY		150	GIUAGPPIOLleAlaphcGLYTHrASnglunGluAInrPserCISnArgrtAgLaGu	169
Db		458	GAAAAACCGCAGTATTAGTGACGATGACGTCACTCATTTCTAAAAACCGTCGCAGTG	517
OY		170	LseSTyr	172
Db		518	TTRAGCTCAC	526
RESULT 32				
LOCUS	U32722/c	13620 bp	DNA	linear BCT 29-MAY-1996
DEFINITION	Haeemophilus influenzae Rd section 37 of 163 of the complete genome			
ACCESSION	U32722.L42023			
VERSION	U32722.1 GI:1573348			
KEYWORDS				
SOURCE ORGANISM	Haeemophilus influenzae Rd. Haeemophilus influenzae Rd Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.			
REFERENCE AUTHORS	1 (bases 1 to 13620) Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A., Merrick,J.M., McKenney,K., Sutton,G.G., Fitzhugh,W., Fields,C.A., Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A., Kelley,J.M., Weidman,J.F., Phillips,C.A., Spitzgs,T., Hedblom,E., Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Sauder,D.M., Brandon,R.C., Fife,L.D., Frithman,J.L., Fuhrmann,J.L., Georgagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M., Smith,H.O. and Venter,J.C. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd			
JOURNAL MEDLINE	Science 269 (5223), 496-512 (1995)			
PUBMED	95350630			
REFERENCE AUTHORS	7542800 2 (bases 1 to 13620) Ratusov,R.L., Mushnegan,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V. Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli			
TITLE	Curr. Biol. 6 (3), 279-291 (1996)			
JOURNAL MEDLINE	96396784			
PUBMED	8605245			
REFERENCE AUTHORS	3 (bases 1 to 13620) White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission			
TITLE	Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
JOURNAL	4 (bases 1 to 13620) White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission			
REFERENCE AUTHORS	Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
REMARK	The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Ratusov et al. We have incorporated their annotation into the /notes fields of the corresponding H. influenzae genes			
REFERENCE	5 (bases 1 to 13620) White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D., Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.			
AUTHORS	Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.			
TITLE	Direct Submission			
JOURNAL	Submitted (28-MAY-1996) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
REMARK COMMENT	The whole genome was shifted by 588 nucleotides for a new start On Sep 30, 1996 this sequence version replaced gi:1221050.			
FEATURES	Location/Qualifiers			
SOURCE	1..13620			
genbank	/organism="Haemophilus influenzae Rd"			
genbank	/db_xref="taxon:71421"			
genbank	complement(61..513)			
genbank	/gene="HI0379"			
genbank	complement(61..513)			


```

/gene="HI0379"
/note="similar to GB:U00096 PID:1788880 percent identity:
63.38; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22037.1"
/db_xref="GI:1573349"
/translation="MKLSKGRVAVTAVIDIALNADGPSVLADISERQHISLYLEQ
LPAKRDGLVKSVRGGGQGLGPSQISVGMITAVNNIHTKCOLGKENCNCEV
ECITHMEDSLKIESFLNETITLAEVLNKNRKRSHDFNNLLVNO"
/complement(566..1291)
/gene="HI0380"
/complement(566..1291)
/note="similar to GB:U00096 PID:1788881 percent identity:
54.43; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22038.1"
/db_xref="GI:1573350"
/translation="MLENIRIVLTETSHSGNIGSAAPAMTGLTQLCLVSPKSVDEQ
SYLSAGAEIVKARVYDSFDEAVDCCPLVTSARLRLQNTLLEPRCAKVVAY
KGTAIYGRERICLTNEELCKCHLNIIPANDPYSILNMAVOLSYELRAFLVQ
NNKNSLSLEKKNPTDOLAFPDYTERIYOSIGFIQONGVAKRKYRAKLEKN
ELNLNGLMSAVEKRIDLTKEDN"
/complement(1405..1477)
/gene="tRNA-Lys-3"
/complement(1405..1477)
/gene="tRNA-Lys-3"
/product="tRNA-Lys-3"
/note="codon recognized: AAG"
/complement(1733..2194)
/gene="HI0381"
/complement(1733..2194)
/gene="HI0381"
/note="similar to SP:P10324 PID:1573351 SP:P10324
GB:M18678 PID:1573351 percent identity: 100.00; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptidoglycan-associated outer membrane
lipoprotein (pal)"
/protein_id="AAC22039.1"
/db_xref="GI:1573351"
/translation="NNKFVKSILVAGSVAALAAACSSNNDAAGNGAAGTEGGYVADL
QQRVNTYVFGFDYDITGEVVOITLDAAVYLNTPAKYLVBSNTDERGTPEPNTALG
ORADAVKGYLACKGVDAKGTGVSIGEEKPAVLGHDEAYTSKNRAVLAY"
/complement(2218..3501)
/gene="HI0382"
/complement(2218..3501)
/note="similar to PID:1685080 percent identity: 96.96;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="colicin tolerance protein (tolB)"
/protein_id="AAC22040.1"
/db_xref="GI:1573352"
/translation="MKLIRKISVFATVAVGSMNFAAGDEVATIDEGVDGARPLAVV
PFVGSAPEDISKIVADDLRNSKRFNPINVSQRPRTSAEYNPEAMSNIGIDATVIG
OVVPSNGSITYTOLDITVAGSGTGETVIMONSYYTNKMLRYGAHVSDEVEFKLTA
IKGAFKRIATVVOVQNGSGSQPYEVRAVDYDGNQPIVNSAOPIMSPAMSPGQRLAY
VSENNKSQLVVDLNSGAKRYVASFOGNGDPAFSPDPSRLAFASPGAGVINTVVG
ANGGTPTQLTSGANGTEPAMPDGNLSILFTSDRSGSPDYVMDASGSGATVGGSG
AOSISADGKTLVINGNNVAKODLTGTVEVISTGFLGSPSLSTNGIMTIITSQGL
GKYLQVADGRRKSLPSGDOVFPFAMSPYLTK"
/complement(3543..4661)
/gene="HI0383"
/complement(3543..4661)
/note="similar to GB:M28232 SP:P19934 PID:148019 GB:U00096

```

```

PID:1651331 percent identity: 42.58; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="outer membrane integrity protein (tolA)"
/protein_id="AAC22041.1"
/db_xref="GI:1573353"
/translation="MQNNRQKKGINAFASITLHFTLFGILLSSLYHVEITMGGEQ
EGDVIAVIVDTGTAOEWGRIOQKGGADQKREPVEERKPPPEEIKKHQOEY
OOREELKROEOOROEIKQOAROEALFEKQOEAKQAQAAPAKLADAEKRR
LAAAKQAEFEKAKAAETIAQAKQAEAKAKLAEAKAVAEKAAKAEAKAKA
AAAKAKADAEKAKATEAKKADQASLDLDFNGDGGGSASGNTNGGSGAA
LGSGDGKVGDDQAGVYIKKEIQRIQLRFLDNFAGKVCRIQIGRQGTILGYKISGD
DICSALSAVAKRKVPAPSDIEYKYSPIIDFDIR"
/complement(4677..5096)
/gene="HI0384"
/complement(4677..5096)
/note="similar to SP:P05829 PID:1128978 GB:U00096
PID:1651330 PID:1786959 percent identity: 61.76;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="colicin transport protein (tolR)"
/protein_id="AAC22042.1"
/db_xref="GI:1573354"
/translation="MARQRKAIKSEINIVPELDVLVLIIFMATAPILISQVOVEL
PDSVQSEVSNEDKVPVILEVAGIKVATISIGSEQEGITTEMTVQLSRQEDDMNT
LFLVNGAKKEVPEYEVYKALNLTLAGIKSVGLMTMPI"
/complement(5168..5854)
/gene="HI0385"
/complement(5168..5854)
/note="similar to SP:P05828 PID:1128978 GB:U00096
PID:1651329 PID:1786958 percent identity: 68.47;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="colicin transport protein (tolD)"
/protein_id="AAC22043.1"
/db_xref="GI:1573355"
/translation="MTAEINFLDLFLKASIVOLVYVILISFIISNAIILORSRIIT
NALKPARTEDEDFMSGDELNKLKESLSNRSGIDGSDPFCVGEKRSRLKQVNDAP
EALIKGTBAMVLANNRITESLNNRVPLATVAVSYSPYIGLGVWGMFAFMALSGA
KOATIQWAPGIAELIATDAIGLFAIPAIVAMVYNLSLRVNAIDQGNFIDEFTTIL
HQAQGRKAPH"
/complement(5871..6281)
/gene="HI0386"
/complement(5871..6281)
/note="similar to SP:P08999 PID:1128977 GB:U00096
PID:1786957 percent identity: 53.08; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22044.1"
/db_xref="GI:1573356"
/translation="MLDNGFSEPVRYVEDTACGVVYHARYLHFEPRAREVLTPLN
FTQOTLLEQDLAVVVTILADYCVAAKLDDLLMVEVESEVKGATILFQRLMRNTL
MLSRATVAVACVDGKMKPVAFPREVNAAPFHHLK"
/complement(6006..8528)
/gene="HI0387"
/complement(6006..8528)

```

Alignment Scores:

	Pred. No.:	Length:
Score:	264.00	13620
Percent Similarity:	68.93%	51
Best Local Similarity:	49.51%	20
Query Match:	30.66%	32
DB:	1	0
Gaps:		0

```

US-09-674-779b-2 (1-172) x U37222 (1-13620)
QY 70 ValTYrPheaspPheaspSeraspGluLeysProGlnAlaAlaAlaLeuaspGlu 89
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2044 GTTATTGCTGGTTGATTAATATGACATTACTGGGATACCTTCACATCTTAAACCCG 1985
QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaAlaValAlaGlyHisHisrAsp 109
    |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1984 CACCGCTGACATTTAAATGCAACGCCAGCTGCTAAAGATTAGTAGAGATACACTGAT 1925
QY 110 GluArgGlySerArgGlyTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1924 GAACGCTGACACACATACACATACGATTAAGCCACGCTGCGACAGATGACATTAA 1865
QY 130 AsnTyrLeuLeuGlyLysGlyLeasngGlnAlaSerValGluIleIleSerPheGlyGlu 149
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1864 GGTATTATTAGCTGCTAAAGCTGTTGATGCTGCTAAATTAGGCACAGTATCTTACGCTGA 1805
QY 150 GluArgProfilLehAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArgAlaGlu 169
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1804 GAAAAACCTGACGATTAAGTCAATGATGAGAGCTGCTATTTCTAAACCGCTGCGACATG 1745
QY 170 LeuSerTyr 172
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1744 TTAGCGTAC 1736

RESULT 33
SME591791 340900 bp DNA linear BCT 05-JUL-2002
LOCUS SME591791
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 10/12.
ACCESSION AL591791 AL591688
VERSION AL591791.1 GI:15075538
KEYWORDS
SOURCE
ORGANISM Sinorhizobium meliloti.
          Sinorhizobium meliloti.
          Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
          Rhizobiaceae; Sinorhizobium.
REFERENCE
AUTHORS Cepeda, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J.,
          Bolstad, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S.,
          Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Leilaure, V., Masuy, D.,
          Pohl, T., Portetlelle, D., Puehler, A., Purnelle, B., Ramsperger, U.,
          Renard, C., Thebaud, P., Vandenbol, M., Weidner, S. and Galibert, F.
          Analysis of the chromosome sequence of the legume symbiont
          Sinorhizobium meliloti strain 1021
          Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
TITLE
JOURNAL
MEDLINE 21396507
PUBMED 11481430
REFERENCE 2 (bases 1 to 340900)
AUTHORS Gouzy, J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
          EU Consortium
COMMENT
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire et Développement UMR6061-CNRS,
Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes,
France, CMC GmbH, Fritz-Arnold-Str. 23, D-78467 Konstanz, Germany,
Universität Bielefeld, Biologie IV (Genetik) universitätsstr 25,
D-33615 Bielefeld, Germany, Unité de Biochimie physiologique,
Université Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unité de Microbiologie, Faculté
des Sciences Agronomiques de Gembloux, Avenue Mareschal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
          Location/Qualifiers
          1..340900
            /organism="Sinorhizobium meliloti"
            /strain="1021"
            /db_xref="taxon:382"
            132..518
            /gene="SMC01981"

FEATURES
source
gene
repeat_region
CDS
132..518
/gene="SMC01981"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence="not_experimental"
/transl_table=11
/product="PUTATIVE CYTOCHROME C TRANSMEMBRANE PROTEIN"
/protein_id="CAC47094.1"
/db_xref="GI:15075539"
/db_xref="SPRMBL:Q92M04"
/translation="MRTALALSPSALISCPVAGLAEQDAEATVFKCATCHVID
KDKKVGPSLQGYGRIGHADPKYQAMIIDAGKGLVMDATLAEYLNNPAXKVG
TKMVPGLKDEETANVATLKQHPK"
/gene="coxM OR SMC01982"
656..1600
/gene="coxM OR SMC01982"
656..1600
/gene="coxM OR SMC01982"
656..1600
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence="not_experimental"
/transl_table=11
/product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
POLYPEPTIDE II TRANSMEMBRANE PROTEIN"
/protein_id="CAC47095.1"
/db_xref="GI:15075540"
/db_xref="SPRMBL:Q92M03"
/translation="MAVVIVLVLAVGSVLEPHLISPMKWPPIASNNVYINTITITPTM
ITGLITAVLFLMAYCVLRFPHRGNTAAYPEPKRRLRGWLATSTTGVAAMAPGLF
VNNQVTVQDASAEVYIGQOMLSFRLPGADGIGTTERDIAPEMTLVNRRDAG
ODDIIEGSELHLPKGVKMLSVLPHDFYVEPRAKMDVPGMTYFWLPTPT
GTFFELCALICGVGHQPMKRGVVDTEEDYQAMLAEOOTFQLSASSETRAVPEKVS
GPSGIATEQGTGASALKRECEGPAATVVASAAQ"
/complement(1457..1513)
/note="Sm-5 OR SMC04646
REPEAT SM-5
Predicted by Homology"
/evidence="not_experimental"
1634..3415
/gene="coxN OR SMC01983"
1634..3415
/gene="coxN OR SMC01983"
1634..3415
/gene="coxN OR SMC01983"
1634..3415
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence="not_experimental"
/transl_table=11
/product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
POLYPEPTIDE I TRANSMEMBRANE PROTEIN"
/protein_id="CAC47096.1"
/db_xref="GI:15075541"
/db_xref="SPRMBL:Q92M02"
/translation="MKVDVRSIGELALPPEVEDVLEIHPHSMTRVYVESODAKTIAI
OYSMTAIGAWVALVSWLIRQLSGPGTFELIDERYIQTITMGMTIVYLLTALT
LGCGNYLIPMGADWVFPYANMLSWIYVILLAVIYLAASEFTPGPLGAWTLTPP

```


Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.
The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58
Science 294 (5550), 2317-2323 (2001)
JOURNAL MEDLINE
21608550
PUBMED
11743193
REFERENCE
AUTHORS
2 (bases 1 to 10906)
Wood, D. W., Setubal, J. C., Kaul, R., Monks, D., Chen, L., Wood, G. E., Zhou, Y., Woo, L., Kitajima, J. P., Okura, V. K., Almeida Jr., N. F., Chen, Y., Boyce, S. D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McCelland, B., Palmeri, A., Raymond, C., Rouse, G., Saenphimachak, C., Wu, Z., Gordon, D., Eisen, J. A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon, Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.
Direct Submission
Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA
FEATURES
source
Location/Qualifiers
1. 10906
/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"
/strain="C58"
/db_xref="taxon:180835"
complement(113..1489)
/gene="mrsa"
/note="Atu3709"
complement(113..1489)
/gene="mrsa"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/transl_table=1
/codon_start=1
/product="phosphoglucosyltransferase/phosphomannomutase"
/protein_id="A144519.1"
/db_xref="GI:17742131"
/translation="MPERGVQARVFGTGRGSGNVEPMTDLMRVGIVGTIF RGRHRRVYIGKTRISGVMLENALVAGTACGADIVLGIPTPAVMALPRLSAD IGVMSIASHNPFSDNGITKLPDPSTKLSDELIEDLDKDIYLAOLAKRAETGRKRV VGDIVYRIEYKRTLPDVTLSGLRTAIDCANGAAKVAAPALWELGAEVYITGEB NGININIECGSTHEALOKKILHEVRADIGALGDADRVIVDERGEIVDGDLMVI ADSAADNTLGGGIVATVMSNGLERFLGDKGLTARTKVRVYVEHRRNPNFVNG GEOSGHTVLSPTGTTGDIVALOVLAKVRSGLVSEVCRKEPEYDOLKKNYISGG KPLNPITVLAIDAESALANNGRLVIRPSTGTEPLIRVAMEGSDSAKEKIVNDIVG V ISSARSA"
complement(1654..3600)
/gene="fthH"
/note="Atu3710"
complement(1654..3600)
/gene="fthH"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/transl_table=1
/codon_start=1
/product="metalloprotease"
/protein_id="A144520.1"
/db_xref="GI:17742132"
/translation="MNPFRNFALMVAILLALLAFSMFQTSPTQSGREIPIYSQIR DVDSGRVDVTVNGTGTENGTAFTQYSPVIDSLMERLQSKNVTIVARESDG SSGFLSYGLTLPMLFLLGVLFPROMOGSGAMGFGSKALLTEAGRYTFDQV AGVDEAKODLEIIEPLRDPQKFORLGGKIPRGVLVGPQGTGKTLARVAGANPV EFTISGSPFVEMFCASRYRDMPEQAKRACPIIFIDEDAGRRHGGIGGNGE REQTLNQLVEMDGEAEAGIITLAAITRPDVLPAIRGRPRQYVNPPIQOE RLKLVHNVPLAPNVDTKITLARGTPGSGDGLMNLVNEALMARKNRVTVQOE DAKDKIMGAERSSAMTEAEKRLTAYHEAGHATLALVAVADPLHAKATLIPRGLA G VMOLPEDGRYSKMYSLVIMMGVRAVELTGEKENTISGASSDIDATKLAGA MYTQMGFSDALGOVAVGENOQEVFLGVSQSVSEATQOTIDTEVRLIDEVYTA RLITDNDHGVAVIAEGLEIETLTGEIKRLALGKRPADLGDSDSGSGSNAVPRAG AKKDPSEAKGDGEAEGMEPPIH"

gene
complement(3755..5086)
/gene="Atu3711"
complement(3755..5086)
/gene="Atu3711"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/transl_table=1
/codon_start=1
/product="cell cycle protein Mesu/cytosine deaminase-related protein"
/protein_id="A144521.1"
/db_xref="GI:17742133"
/translation="MPVDARPNKTYAPLAASAAYESLRKPHIVALLTSGSDSG LLLAQEKRAADDGLRLSAVTVHARLPESADARKVAALICELGIPHTTRMDPK PASGISEASRLARLADVAREIGADLVSGHTIGQRETVAMARASGSNDGLS GMADAVYDALIMRPEFLACERQIVDYVSSRSHWLDDPVENIYERYVRQTL HSPALDSTAVROMLSERTAVFLRERAGVFAALRLADDVNPDLPEFRGLAL IALGCRSPFAASSMEVRLFKETGTGRTVGRVLLDRRDDLYIFRQRNLPQL IESSQAVMDRPLRYKNSDPRPIVAVAGRGCTAQAALFPASGVKXPMGMLPQI AAAAGWVSTPEVAITPRLAFDLPFLPDLLELANALNLFGRATYQPPY"
complement(5111..6100)
/gene="Atu3712"
complement(5111..6100)
/gene="Atu3712"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/transl_table=1
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="A144522.1"
/db_xref="GI:17742134"
/translation="MAGMLGLACGSLGTVASPLRGAGSGFTINSOOPITRYVA NEAYVQOLEQDIRQLNRIEEMSFQLQWETIRKAOEDNEFPQLEBGGKAPRA APKRAKASVPNPAPPTDDVATLIEPPQGAASVSPSSASTALAGETLSTIL DSKGMPVGTILNOGANNSSGSLPGVTTGNGRKQDPVNTALITSEGDLYQAYGHVLS GPKYLAEGFOOYLOGYPKGTAKADASFMLEAGYISQKFEAKTELNQHOTGKSP KAPEMLKMGISLAIDNTEACATLRVPRYRNASTVLSKVASQKRLSC"
complement(6274..6807)
/gene="pala"
/note="Atu3713"
complement(6274..6807)
/gene="pala"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/transl_table=1
/codon_start=1
/product="omp16 protein"
/protein_id="A144523.1"
/db_xref="GI:17742135"
/translation="MSRTNISALSPMOKLARNPVIAMTLALAGCANKNMPSAG ELGLGAGSATPGSQDDPTVYNGDRILFPDQSTSRADAOOTLROAOWLSRYPNVAI TVEGHADERTREYNIALGARRAATRFLLASQGVASRMKTTISGKREKPAVVCDDIS CWSQNRRAVTVGGAGM"
complement(7018..8325)
/gene="tolB"
/note="Atu3714"
complement(7018..8325)
/gene="tolB"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/transl_table=1
/codon_start=1
/product="tolB protein"
/protein_id="A144524.1"
/db_xref="GI:17742136"
/translation="MKIFSPILRVLAIALMSVFSAFAFARVEININKNVEPLPIAI TQMSSDGIGAOVSAVIADLQSGFLAPVNNAFIERISNDDQPREDEKVMIAQA LVYGRVTRSDGLRAERLMDPTFPAGQOLSGOQFTPEPMNRVAHIIADAIYERVTG EKGTPTRVYVSESGPKARKRQSLINDQDFENRLTNSDYLTPRSPNROEVY TYSFEQDQPRVILQLETQKRVNGFPMATSPRSPDQCVINISLOODCANIYTM DLSRTYVPLRLNTAALDPAVSYSPGQVAFESDGRQITVWADNSGQQRSPFG DGSYTPVMSRBDLLAFTRKQSGKFSIGVAKTDGSGERLITSGFHNBPVAPMGRLV MFRQWAGAGGQVLSIDLTYGNEQLIKPTPASPAMSPILD"

[illegible]


```

Oy      149 glugluargproleialaPheGlyThrasnclugluAlaTrpsergluAsnArgAla 168
Db      580 AAGGAAAGCCGCGCGATTGGACACACATTTCCTGCTGCTCCACAGACCCTCGGCGCT 639

RESULT 36
LOCUS   AE012837                      10643 bp    DNA        linear    BCT 01-Jul-2002
DEFINITION  Chlorobium tepidum TLS section 58 of 194 of the complete genome.
ACCESSION  AE012837 AE006470
VERSION    AE012837.1 GI:21646572
KEYWORDS
SOURCE
ORGANISM  Chlorobium tepidum TLS.
           Chlorobium tepidum TLS
           Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
           Chlorobium.
REFERENCE  1 (bases 1 to 10643)
AUTHORS   Eelsen,J.A., Nelson,K.E., Paulsen,I.T., Heidelberg,J.F., Wu,M.,
           Dodson,R.J., Debroy,R., Gwin,M.L., Nelson,W.C., Haft,D.H.,
           Hickey,E.K., Peterson,J.D., Durkin,A.S., Kolonay,J.L., Yang,F.,
           Holt,I., Umayam,L.A., Mason,T., Brenner,M., Shea,T.P., Parksey,D.,
           Niernan,W.C., Feldblyum,T.V., Hansen,C.L., Craven,M.B., Radune,D.,
           Vamathevan,J., Khouri,H., White,O., Gruber,T.M., Ketchum,K.A.,
           Venter,J.C., Tettelin,H., Bryant,D.A. and Fraser,C.M.
           The complete genome sequence of Chlorobium tepidum TLS, a
           photosynthetic, anaerobic, green-sulfur bacterium
           Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)
           2 (bases 1 to 10643)
           Eelsen,J.A., Nelson,K.E., Paulsen,I.T., Heidelberg,J.F., Wu,M.,
           Dodson,R.J., Debroy,R., Gwin,M.L., Nelson,W.C., Haft,D.H.,
           Hickey,E.K., Peterson,J.D., Durkin,A.S., Kolonay,J.L., Yang,F.,
           Holt,I., Umayam,L.A., Mason,T., Brenner,M., Shea,T.P., Parksey,D.,
           Niernan,W.C., Feldblyum,T.V., Hansen,C.L., Craven,M.B., Radune,D.,
           Vamathevan,J., Khouri,H., White,O., Gruber,T.M., Ketchum,K.A.,
           Venter,J.C., Tettelin,H., Bryant,D.A. and Fraser,C.M.
           Direct Submission
           Submitted (30-APR-2002) The Institute for Genomic Research, 9712
           Medical Center Dr, Rockville, MD 20850, USA
           unpublished

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

REMARK
FEATURES
SOURCE

Location/Qualifiers
1..10643
/organism="Chlorobium tepidum TLS"
/strain="TLS"
/db_xref="taxon:194439"
117..1421
/gene="CT0636"
117..1421
/gene="CT0636"
/feature="similar to GB:M28232 SP:P19935 PID:148020 GB:U00096
PID:1651332; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ToB protein, putative"
/protein_id="AAW71875.1"
/db_xref="GI:21646573"
/transl_table=11
/translation="MRSTRNSFRCICIMLFGLFVFTLRAEVEGYIAIRREGASRI
AAVLDKTSADGKQREKARSLDVTINKGDFGLNLPAIPINRGQNGGKIPASIA
SVGGDIYAGSGVTRKSRGRLVLEHMYDSSGSLATATYGESOLRAICRCAQVE
LTLGKRSVGTIRIVFANRTGNKEIYCDPGEENYVOLNRSISLTLPVSPDGITA
WTDYISGRPNLYIKNIATGAKYSVNKGVCISPARPGNPTLVTLSTGDDDLIR
ADGVTRRLTKGGGIDVSPTPSGSKAFVSTROGCPPIFDAMNSGGVRLITSGI
YINPSPSPNGKILXSSQKSGEIRIITSINVDGSLDITSGSNNEHPSPSGMS
IYPSSTRDGRRLRYNNADSGNORPLNNGBQOPSWSVSK"
1614..1757
/gene="CT0637"
1614..1757
/gene="CT0637"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW71876.1"

gene
CDS

/db_xref="GI:21646574"
/transl_table=11
/translation="MQVGMSFGADQAMRTTISYSEKPFDLGHDETAGSKRWAAH
VER"
1834..2346
/gene="CT0638"
1834..2346
/gene="CT0638"
/feature="identified by match to PPM protein family HMM
PF00691"
/codon_start=1
/transl_table=11
/product="peptidoglycan-associated lipoprotein"
/protein_id="AAW71877.1"
/db_xref="GI:21646575"
/transl_table=11
/translation="MRRLTGLIGAMTILAGCSSKAVSTDETSPAGSGGGGTG
AGAGVVEDIGGGKAGSLITDIFDFOSALSSPAOLNMAAMQKNPTSVITIE
KRRDHFVYK"
complement(2384..2482)
/gene="CT0639"
complement(2384..2482)
/gene="CT0639"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW71878.1"
/db_xref="GI:21646576"
/transl_table=11
/translation="MQKRWQEPTRKDVFWSSWIDIMFNNTGGMA"
2455..3243
/gene="CT0640"
2455..3243
/gene="CT0640"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW71879.1"
/db_xref="GI:21646577"
/transl_table=11
/translation="WPAIFLISACASQNDLSYVOGVSQLKQESTVIKQSGAGSSEM
TOYREIASLNGKIDELQYDRTSKRLDMDSILVRKVDLENRIRARIEQYLGIST
GKDSILVRKVPPEPPSSSSNSGTSQAQSEKASATGALLSEGLIKKRGDYG
ARESEFNAPMTGNPKPVADAQOFIAETVYEMKYEKRAILEYOTVIRATYKSPRRPA
LYKQGLSPAKIGDEANAKARYKDVNLNLPQSEPAKLNQKNLDRK"
complement(3366..3797)
/gene="CT0641"
complement(3366..3797)
/gene="CT0641"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="photosystem P840 reaction center protein PscD"
/protein_id="AAW71880.1"
/db_xref="GI:21646578"
/transl_table=11
/translation="MOPQLSPRQASNOVRKAVSGPMSGNVHAKEYFTTSKKRQD
GKIDLEIYVPSAGRGKLSPTPMIRLUDGELIETITDIDALDNKKETIDMENRY
IDPFRGKWTWRELPIVFIHGKGLCVLHNKITYTDQFR"
4141..4590
/gene="CT0642"
4141..4590
/gene="CT0642"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW71881.1"
/db_xref="GI:21646579"
/transl_table=11
/translation="MHRKERALEFSNRCSOAVFAAPROTVDLQASRLATATFG
GGVAGSGGMCAGATGALIMLSRKYGGVDELYNKKITYELRQRTIEERKMSGAR
CESITLGCIGEPENLQKARLKLFEYVSAVAFTASDILEMLCAG"
complement(4720..6621)

```

/gene="dnak"
 /note="CT0643"
 complement(4720..6621)
 /gene="dnak"
 /note="identified by match to PFAM protein family HMM PF00012"
 /codon_start=1
 /transl_table=11
 /product="Dnak protein"
 /protein_id="AAW71882.1"
 /db_xref="GI:21646582"
 /translation="MGKIGIDIGTNSCAVMGCTOPTVIENSSEGRPTSPMAVATK
 TGERLVGAARKRAVTPNPKTITSIKREGRKRYDPNPKTIASDVNVEGGYARKI
 GDRYSPQETISAMTLOKOTADFLGEKTEVATVPATFNDAGQATKDKAGTAGL
 EVKRIINEPTAAALAYGLDKKRENERVAVFDLGGFTDLSLEGGVEVSTDDT
 HLGDDPDQVITINYLADFEKQEGIDLRDADALQRLKRAAEKAKIELSRDTEINT
 PFTAQOEGEKHLVINLTKAFKFAAMALFDLFEPCRAIKSKRFDIEIDIEVYVG
 GSTRIPVQALVKEFEFGKEPNKSVNPDEVVAIGALOGGVLDGVDVLLVTPSL
 GFTLGVMTKLEANTTPTPKROEIESTADNOSTVEVHVLGGDEPMASDKTKGFE
 HSDTPAPRGVQIETPTDIDANGILNLSAKRKATGKQSTIKESGKLTAEIEKM
 KEDAKAAAEADQKREIETLKSADSLITSEKQLELGDLPADKKAIESALEKIK
 EAKSGRVDAIKRPAMDELKRVMSDASNLVGPAPPEPPTNGHAGSGSGGGAANA
 YEVIDGDDK"
 complement(6701..7099)
 /gene="CT0644"
 complement(6701..7099)
 /note="identified by match to PFAM protein family HMM PF00011"
 /codon_start=1
 /transl_table=11
 /product="heat shock protein, Hsp20 family"
 /protein_id="AAW71883.1"
 /db_xref="GI:21646581"
 /translation="MLMKIAKDPRLFDIDWSSQMAVAFSEKVIDISEDNAYHDAE
 LPGAKEQIALNIEDVLTITKAERTHEEKKNYHVERTYGSFSRSENGEIIQOE
 HIGATVNGVLIHTLPKTOPAKTKKEIPIN"
 7118..7228
 /gene="CT0645"
 7118..7228
 /note="identified by glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAW71884.1"
 /db_xref="GI:21646582"
 /translation="MFVFSRFGMLRSILTSVETSKDVPKRFHFDLCV"
 complement(7467..7856)
 /gene="CT0646"
 complement(7467..7856)
 /gene="CT0646"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator, ArsR family"
 /protein_id="AAW71885.1"
 /db_xref="GI:21646583"
 /translation="MKSMTTTEEVKKWQFNPMDMLDAVSNRKLSPMKILR
 ALCDREHTVQETVKEGASQANISKHIALMDNVDVNRKRGKCYRIKIDDDSVYAC
 FLISKSVENLQDRISLWIKVNTWLT"
 complement(8023..8250)

Alignment Scores:

Pred. No.: 5,82e-17
 Score: 259.50
 Percent Similarity: 56.41%
 Best Local Similarity: 35.90%
 Query Match: 30.14%

DB: 1
 Gaps: 2

US-09-674-779b-2 (1-172) x AE012837 (1-10643)

QY 15 SerValLeuThrPheMetThrGlyCysAlaAnlySerThrSerGlnValMetValAla 34

Db 1864 GCGGCATGATCATTTTGGCAGGCTGTCTCAAAAAGTGA-----GTATCAACAGAT 1917
 QY 35 ProAnaLapProthGlyTyrThrGlyValIleYrThrGlyVal-----AlaPro 51
 Db 1918 GAACCTCCAGACGCCGTTTATGATCGGAGTGGAGCGGCACTGACGACGAGCAGGA 1977
 QY 52 LeuValAspAnaSpGluThrValIleAlaLeuAlaSerIysLeuProSerLeuValTyr 71
 Db 1978 GTGAGCTGTGAAGATATCGCCCAAGGGGCAAGGCCGTTCAATTTATGGATATTTTC 2037
 QY 72 PheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGluAla 91
 Db 2038 TTGCAATTCGACAGTTCGCGCACTAGCTCCAGACACAGAACTCAATCAAAAGCC 2097
 QY 92 GlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArg 111
 Db 2098 GCCTGGATGACAGAAAACCTACTCTGCGGTATCATTTAGGACATTCGGATGACGCG 2157
 QY 112 GlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyr 131
 Db 2158 GGGACCGATGACTACACATCGCGCTGTAGCGTGTGTAAGCAGCAGAAATGAT 2217
 QY 132 LeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArg 151
 Db 2218 CTGTTTAACTTGAGTGAAGCGAGCGCGCTTCTTACCGTCAGTTAGCGGAGAAAG 2277
 QY 152 ProIleAlaPheGlyThrAsnGluGluIleArgSerGlnAsnArgArg 167
 Db 2278 CCGTTTGACCCCGGTCATACAGAGAGCTTGCGCAAGAACAGAGA 2325

RESULT 37
 AE004259/c
 LOCUS
 DEFINITION
 Vibrio cholerae chromosome I, section 167 of 251 of the complete
 chromosome
 AE004259 AE003852
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Vibrio cholerae.

REFERENCE
 1 (bases 1 to 11914)
 Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L.,
 Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
 Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
 Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,
 Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D.,
 Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
 Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
 Direct Submision
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

TITLE
 JOURNAL
 FEATURES
 source
 1..11914
 /organism="Vibrio cholerae"
 /strain="N16961"
 /serotype="O1"
 /db_xref="taxon:666"
 /chromosome="I"
 /note="biotype: El Tor"

ACCESSION	U16849
VERSION	U16849.1
KEYWORDS	GI:841125
SOURCE	Pasteurella multocida.
ORGANISM	Pasteurella multocida Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Pasteurella.
REFERENCE	1 (bases 1 to 453)
AUTHORS	Kasten,R.W., Hansen,L.M., Hinojoza,J., Bieber,D., Ruehl,W.W. and Hirsch,D.C.
TITLE	Pasteurella multocida produces a protein with homology to the p6 outer membrane protein of Haemophilus influenzae
JOURNAL	Infect. Immun. 63 (3), 989-993 (1995)
MEDLINE	95172751
PUBMED	7868272
REFERENCE	2 (bases 1 to 453)
AUTHORS	Kasten,R.W.
TITLE	Direct Submission
JOURNAL	Submitted (03-NOV-1994) Rickie W. Kasten, Vet. Med. PMI, University of California, Davis, CA 95616, USA
FEATURES	Location/Qualifiers
source	1..453
	/organism="Pasteurella multocida"
	/strain="T16"
	/specific_host="Meleagris gallopavo"
	/db_xref="taxon:747"
	/clone="F168N"
	1..453
CDS	/note="outer membrane protein, similar to Haemophilus influenzae outer membrane protein P6, Swiss-Prot Accession Number P10324"
	/codon_start=1
	/transl_table=1
	/protein_id="AAI79373.1"
	/db_xref="GI:841126"
	/translation="MKRTKVLIVAGSVAVLAACGSSKKDESQMGFGYSVDLQGR YNYRYRGDRKNIGERYVOIIDAHAFLEMTAPRVVVEGFTDERGTPEVNIALGORR ADAKHTLSAAGVADGVSITYGEERPAVLGHDEAYSKNRRAVLAY"
BASE COUNT	148 a 79 c 98 g 128 t
ORIGIN	
Alignment Scores:	
Pred. No.: 1	1.46e-18
Score: 258.00	Length: 453
Percent Similarity: 66.9%	Matches: 50
Best Local Similarity: 48.54%	Conservative: 19
Query Match: 29.97%	Mismatches: 34
DB: 1	Indels: 0
	Gaps: 0
US-09-674-779B-2 (1-172) x PMU16849 (1-453)	
OY	70 ValtyrPheaspPheaspSeraspGluileysProGlnAlalaaleLeuaspGlu 89
Db	: :
	142 GTGATTTCGGCTCGATTAATAACAATGCAAGTGGAATATGTACAATTTAGATGCA 201
OY	90 GlnAlaGlnPheleutrrThrAsnGlnThrAlaArgValleuAlaAglyHisThasp 109
Db	:: :: :: :: :: ::
	202 CATCGGCCTTTCTTAATGCAACACCACGCAACGAAAGTCGTGTGAAGGTAAACACCAT 261
OY	110 GluArgGlySerArgGluTyrrAsmetSerLeucGlyGluArgArgAlaValAlaValarg 129
Db	:: :: :: :: :: :: :: :: :: :: ::
	262 GAAGCGCGTACACACAGAATATTAACATGCAATGATAGTCAACGTCGTGACAGATGCAAGTAAA 321
OY	130 AsnTyrlleuLeugLytsGlylleasGlnAlaSerValdlulelleeserPieglyGlu 149
Db	:: :: :: :: :: :: :: :: :: :: ::
	322 CATTATTATACAGCAAGAGGTGTACAAAGCTGTCGACTCAATCAACAGACTTCTTCAGGTGA 381
OY	150 GluArgProIleAlaPheGlyThrAsngLuGlualatrpserGlnAsnarGArgAlaGlu 169
Db	:: :: :: :: :: :: :: :: :: :: ::
	382 GAGAACCTCCTGCTGTAGCTCACAGATGAGCAGACTTACTCTAAAAACGTCGCTGCTG 441
OY	170 leuSerTyr 172
	::

DB	442	TTAGCATAC	450
RESULT 40			
AE006136/c			
LOCUS	AE006136	11061 bp	DNA linear BCT 08-MAR-2001
DEFINITION	Pasteurella multocida PM70 section 103 of 204 of the complete genome.		
ACCESSION	AE006136	AE004439	
VERSION	AE006136.1	GI:12721286	
KEYWORDS			
SOURCE			
ORGANISM	Pasteurella multocida.		
REFERENCE	Pasteurella multocida		
AUTHORS	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; 1 (bases 1 to 11061)		
TITLE	May, B.J., Zhang, Q., Li, L.L., Paustian, M.L., Whittam, T.S. and Kapur, V.		
JOURNAL	Complete genomic sequence of Pasteurella multocida, PM70		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)		
PUBMED	21145866		
REFERENCE	11248100		
AUTHORS	2 (bases 1 to 11061)		
TITLE	Zhang, Q. and Kapur, V.		
JOURNAL	Direct Submission		
PMID	Submitted (24-OCT-2000)		
LOCUS	University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA		
FEATURES	Location/Qualifiers		
SOURCE	1..11061		
gene	/organism="Pasteurella multocida"		
CDS	/strain="PM70"		
gene	/db_xref="taxon:747"		
CDS	complement(67..357)		
gene	/gene="PM0962"		
CDS	complement(67..357)		
gene	/gene="PM0962"		
CDS	/codon_start=1		
gene	/transl_table=11		
CDS	/product="unknown"		
gene	/protein_id="AAK03046.1"		
CDS	/db_xref="GI:12721287"		
gene	/translation="MIFYDKGMSIPVLAMLMFSGIFPSTQOMLSYQORAIYIQ HVGALQIAQNRQROPLGACENNVRKNDLVFQIQCTSHKVNSTSLAETSL"		
CDS	complement(341..988)		
gene	/gene="PM0963"		
CDS	complement(341..988)		
gene	/gene="PM0963"		
CDS	/codon_start=1		
gene	/transl_table=11		
CDS	/product="unknown"		
gene	/protein_id="AAK03047.1"		
CDS	/db_xref="GI:12721288"		
gene	/translation="MIFYDKGMSIPVLAMLMFSGIFPSTQOMLSYQORAIYIQ VQOMLVQRONEKEKQHLCKHISTSENNQVTFRLVEGKANYVMCICPTLRQAPR QANIKKLAFLIQRREYRGLESHQFSVOHAPYFYWPTDKQSWTLTONIEATITAEEDL HINGKRIKGTITVIGGRILTEPVSQVSYHKATYAVWYTLRDMQLAEKSWHDFLL"		
CDS	complement(985..1698)		
gene	/gene="PM0964"		
CDS	complement(985..1698)		
gene	/gene="PM0964"		
CDS	/codon_start=1		
gene	/transl_table=11		
CDS	/product="unknown"		
gene	/protein_id="AAK03048.1"		
CDS	/db_xref="GI:12721289"		
gene	/translation="MLFRGYTVLSLMSISLSLLLVIAFYVDVQRONRAIHFMQQLDAELETYQLMGKDIRAGFRAPETYNKVFSLSESTQASIQLSIHGVAVSDCI LPEFYIDMNGCIGGKFKETCIKDGNNSTSHLERELFGRLNNGMELPRTKSAVNS RKLAEKRTYLOOSACHSGGWVDLDDQGYITISDLSFHMIMKRPALFETLRKTIKRP HFOYETTVAPLPLNGEKO"		
CDS	complement(1702..2229)		
gene	/gene="PM0965"		
CDS	complement(1702..2229)		

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:20 ; Search time 215 Seconds
(without alignments)
1801.599 Million cell updates/sec

Title: US-09-674-779b-2
Perfect score: 861
Sequence: 1 MMHLHQAIAAAALSVLTFM.....IAFGTNEBANSQNRRAELSY 172

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Command line parameters:
-MODEL-frame+ .p2n.model -DEV=xlp
-O=/cqr2.1/USPTO.spool/US09674779.rng -MINMATCH=0.1 -LOOPEL=0
-DB=N.Geneseq.101002 -QPM=fastap -SUFFIX=p2n.rng -TRANS=human40.cdi
-LIST=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -MINLEN=0 -ALIGN=40
-MODE=LOCAL -OUTFM=ptc -THR_SCORE=500 -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-USBR=US09674779.ecgn.1.1.208.e_rnat.06072003.121613.9795 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N_Geneseq.101002.*
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	861	100.0	519	21 AA240351	M. catarrhalis BAS
2	861	100.0	92407	22 AAF28549	Genomic fragment #
3	855	99.3	519	21 AA240352	M. catarrhalis BAS
4	850	98.7	519	21 AA240353	M. catarrhalis BAS
5	844	98.0	519	21 AA240354	M. catarrhalis BAS
6	269.5	31.3	866	11 AA006089	Sequence encoding
7	264	30.7	737	11 AA003869	H. influenzae DNA f
8	264	30.7	867	9 AA081194	16600 dalton outer
9	264	30.7	1019	15 AA045440	Fragment encoding
10	264	30.7	1830121	17 AA742063	Haemophilus influe
11	259	30.1	737	9 AAN80226	Sequence of Haemop
12	215.5	25.0	540	18 AAT67789	H. pylori outer me
13	215.5	25.0	540	18 AAT77469	H. pylori outer me
14	215.5	25.0	540	22 AAF25593	H. pylori HPS14 e
15	215.5	25.0	561	18 AAT68048	H. pylori outer me
16	211.5	24.6	770	20 AAV90653	Nucleotide sequenc
17	211.5	24.6	1236	20 AAV90843	Nucleotide sequenc
18	211.5	24.6	1610	20 AAV90545	Nucleotide sequenc
19	205.5	23.9	435	18 AAT74194	H. pylori Omp22 ge
20	204.5	23.8	579	24 ABL91202	Chlamydia pneumoni
21	204.5	23.8	779	22 AAD08593	Chlamydia pneumoni
22	204.5	23.8	1230025	20 AAX91990	Nucleotide sequenc
23	196.5	22.8	1038602	20 AAX01425	Complete genome se
24	180.5	21.0	675	22 AAF30043	Moraxella catarrha
25	180.5	21.0	31940	22 AAF28526	Genomic fragment #
26	176.5	20.5	1830121	17 AAT42063	Haemophilus influe
27	175.5	20.4	645	17 AAT32600	P. aeruginosa OPrF
28	172	20.0	4274	14 AAO54144	Sequence of Plasmid
29	171.5	19.9	486	17 AAT32599	P. aeruginosa OPrF
30	171.5	19.9	681	17 AAT34419	P. aeruginosa OPrI
31	171.5	19.9	1253	9 AAN82023	Outer membrane pro
32	170	19.7	1319	21 AAC38558	Actinobacillus ple
33	169	19.6	1110	21 AAC79664	Virulence gene #71
34	165.5	19.2	1720	16 AAT60916	Non-typable Haemop
35	165	19.2	898	24 AAT46503	M. catarrhalis MCA1
36	165	19.2	96109	22 AAT28548	Genomic fragment #
37	161.5	18.8	1035	24 ABA91419	Haemophilus paraga
38	161.5	18.8	1035	24 ABA91422	Haemophilus paraga
39	161.5	18.8	1035	24 ABA91423	Haemophilus paraga
40	161.5	18.8	1035	24 ABA91425	Haemophilus paraga
41	161	18.7	1026	24 ABQ90179	Haemophilus paraga
42	160	18.6	981	22 AAF23499	M. capsulatus gene
43	158.5	18.4	1035	24 ABA91417	C. coli Cadr DNA.
44	158.5	18.4	1035	24 ABA91418	Haemophilus paraga
45	158.5	18.4	1035	24 ABA91420	Haemophilus paraga
46	158.5	18.4	1035	24 ABA91426	Haemophilus paraga
47	155.5	18.1	1026	24 ABA91421	Haemophilus paraga
48	155.5	18.1	1026	24 ABA91424	Haemophilus paraga
49	153.5	17.2	1026	24 ABA91427	Haemophilus paraga
50	148.5	17.2	6477	11 ABA81540	N. meningitidis pa
51	148.5	17.2	6477	11 ABA81540	N. meningitidis pa
52	148.5	17.2	349980	21 AAF21607	Plasmid pInf 4-49.
53	148.5	17.2	1437668	21 AAN81490	Neisseria meningit
54	147	17.2	66788	23 AAS59515	N. meningitidis B
55	145	16.8	861	22 AAF23500	Propionibacterium
56	145	16.8	861	22 AAF23501	C. jejuni cadf-M129
57	144	16.7	669	20 AAX91705	C. jejuni cadf-F380
58	144	16.7	690	20 AAX91580	Porphyromonas ging
59	143.5	16.7	5579	11 AAO02032	Porphyromonas ging
60	143	16.6	4403765	22 AAT96683	Plasmid pHS 164.
61	143	16.6	4411529	22 AAT96682	Myobacterium tube
62	142.5	16.6	1095	21 AAC79663	Virulence gene #70
63	142	16.5	2016	20 AAX91718	Porphyromonas ging
64	142	16.5	2037	20 AAX91591	Porphyromonas ging
65	138.5	15.1	1922	21 AAA38557	Actinobacillus pie
66	136.5	15.9	411	24 ABA93312	M. catarrhalis MCA1
67	136.5	15.9	691	24 AAT46497	M. catarrhalis MCA1

68	136.5	15.9	730	17	AAT28425	K. pneumoniae dete
69	136.5	15.9	730	22	ABA76835	Klebsiella pneumon
70	136.5	15.9	1008	16	AAT03490	K. pneumoniae p40 g
71	136.5	15.9	1008	17	AAT13607	Klebsiella pneumon
72	136.5	15.9	1008	18	AAV13867	Gene coding for P4
73	136.5	15.9	1032	20	AAZ30477	K. pneumoniae OmpA
74	136.5	15.9	1032	22	AAH78461	Nucleotide sequenc
75	136.5	15.9	1032	22	AAF90077	Nucleotide sequenc
76	136.5	15.9	1032	22	AAF80152	Nucleotide sequenc
77	136.5	15.9	1035	18	AAV13866	Gene coding for LP
78	136.5	15.9	1035	21	AAV13868	DNA encoding a P40
79	136.5	15.9	1035	21	AAV15081	DNA encoding a K1e
80	136.5	15.9	1035	21	AAV15036	CDNA encoding a P4
81	136.5	15.9	1035	21	AAAG3956	DNA encoding an ou
82	136.5	15.9	1035	21	AAAG3961	DNA encoding an ou
83	136.5	15.9	1035	21	AAAI5498	DNA encoding a K1e
84	136.5	15.9	1035	22	AAH74731	Nucleotide sequenc
85	136.5	15.9	1356	22	AAF80153	Nucleotide sequenc
86	136.5	15.8	1727	16	AAQ91871	Branhamella catarr
87	135.5	15.7	640681	24	ABA92787	Buchnera sp. genom
88	130.5	15.2	2251	22	AAC85140	R. anatispefifer o
89	130.5	15.2	2251	22	AAF28554	Genomic fragment #
90	129	15.0	792	24	ABQ90135	M. capsulatus gene
91	122.5	14.2	1650	22	ABA89117	Escherichia coli p
92	122.5	14.2	28989	22	ABAB89112	Escherichia coli p
93	117	13.6	1740	23	ABL53296	2-ketoadonate red
94	117	13.6	1788	23	AA889925	DNA encoding novel
95	117	13.6	1788	23	AA889841	DNA encoding novel
96	117	13.6	1788	23	AA889919	DNA encoding novel
97	117	13.6	1788	23	AA889924	DNA encoding novel
98	117	13.6	2377	23	AA877370	DNA encoding novel
99	117	13.6	21948	20	AA820562	Polynucleotide seq
100	115	13.4	242	20	AAV90680	Nucleotide sequenc

ALIGNMENTS

RESULT 1
ID AA240351 strand; DNA; 519 BP.

XX AA240351;

DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein coding sequence #1.

XX BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
KW upper respiratory tract infection; middle ear infection; therapy; ss.

XX Moraxella catarrhalis.

PN W09957277-A2.

PD 11-NOV-1999.

PF 03-MAY-1999; 99WO-EP03038.

PR 06-MAY-1998; 98GB-0009683.

PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Ruelle J;

DR WPI: 2000-062148/05.

DR P-PSDB: AAY55089.

PT Novel BASB019 polynucleotides and polypeptides from Moraxella
catarrhalis used to prepare vaccines against bacterial infections

PS Claim 10; Fig 2; 101p; English.

XX This sequence encodes a Moraxella catarrhalis BASB019 protein of the
CC invention. The sequences can be used for diagnosis of disease, staging of
CC disease, or determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and for
CC screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB019 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies, and as a
CC target for the screening of antimicrobial drugs. The polypeptides can
CC also be used in vaccine formulations, and to identify agonists and
CC antagonists. The polypeptides, antibodies, agonists and antagonists
CC (which are bacteriostatic) are used for the treatment and prevention of
CC diseases including bacterial infection, otitis media in infants and
CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The frequency of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.

XX SQ Sequence 519 BP; 139 A; 108 C; 122 G; 150 T; 0 other;

Alignment Scores:

Pred. No.: 7,59e-96 Length: 519
Score: 861.00 Matches: 172
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-674-779b-2 (1-172) x AA240351 (1-519)

QY	1	MetMetLeuHisIleGlnIleAlaAlaAlaAlaLeuSerValLeuThrPheMet	20
DB	1	ATGATGTTACATATTCATAAATTCGCCGCGCTTATCGTATCATCTTTATG	60
QY	21	ThrGlyCysAlaAsnIysSerThrSerGlnValMetValAlaProAsnAlaProThrGly	40
DB	61	ACAGGCTGTGCCATATAATCAACAGTCAAGTTATGGTCTCTATGACCCACAGCT	120
QY	41	TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValIys	60
DB	121	TACACGTGGGGTTATCTATCTGTTGACCTTGTGTGATATGATGACACCGTTAAG	180
QY	61	AlaLeuAlaSerIysLeuProSerLeuValTyrPheAspPheAspSerAspIuLeuIys	80
DB	181	GCTCTGGCAAGCAAGTACCACTTGGTTGTTATTTTGGACTTGATCTGATAGATTAA	240
QY	81	ProGlnAlaAlaAlaIleLeuAspGluGlnIleAlaGlnPheLeuThrThrAsnGlnIleAla	100
DB	241	CCGCAAGCTGCTGCCATCTTACGACGACCAACAAATTTTAAACCAACCAATCAACAGT	300
QY	101	ArgValLeuValAlaGlyIleThrAspGluArgGlySerArgGluTyrAsnMetSerLeu	120
DB	301	CGTGTGTTGGTTCAGAGTCATACCGATGAGCGGTGATGCGATATATATGTCACAG	360
QY	121	GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyIleAsnGlnAla	140
DB	361	GGGGAACGCCGTGGCGGTGCGGATTCCTTGGTAAAGCATTAATCAAGCC	420
QY	141	SerValGlnIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu	160
DB	421	AGCGTTGAGATTATACGTTTGGTGAAGACGCCCTATGCGATTGGCACAAATGACAA	480

QY 161 AlaTSPserGlnAsnArgAlaGluLeuSerTyr 172
 Db 481 GCATGTCACAAATCGTCGTGCAACTGCTTAT 516

RESULT 2

AA28549/c
 AAF28549 standard; DNA: 92407 BP.

XX AAF28549;

DT 04-APR-2001 (first entry)

DE Genomic fragment #36.

KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KM bronchopulmonary; endocarditis; meningitis; ss.

OS Moraxella catarrhalis.

PN WO200078968-A2.

PD 28-DEC-2000.

PF 16-JUN-2000; 2000WO-US16649.

PR 18-JUN-1999; 99US-0140121.

PA (INCY-) INCYTE GENOMICS INC.

PI Lagace RE, Patterson C, Berg KL;

DR WPI; 2001-041427/05.

PT Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -

PS Claim 1; Page 369-391; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28549-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.

SQ Sequence 92407 BP; 26788 A; 17581 C; 20150 G; 27888 T; 0 other;

Alignment Scores:

Pred. No.:	1.03e-92	Length:	92407
Score:	861.00	Matches:	172
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-674-779B-2 (1-172) x AAF28549 (1-92407)

QY 1 MetMetLeuHisIleGlnIleAlaIleAlaIleAlaLeuSerValIleuThrPheMet 20

Db 26057 AAGAGGTATACATATCAATATCCGCCGCTGCGCGCTTATCGTACTAATTATG 25998

QY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40

Db 25997 ACAGGCTGCGCAATAAACAACAAGTCAAGTTATGCTGCTCAATGACACCCACAGT 25938

QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60

Db 25937 TACACTGGGGTTATCTATACATCGTGTGCACCTTTGGTAGATGAATGATGACACCGTTAAG 25878

QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80

Db 25877 GCTCTGGCAGACACTACCCAGTTTGGTTATTTTGACTTTGATTCATGATGATTTAA 25818

QY 81 ProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100

Db 25817 CCGCAAGCTGCTGCATCTTAGACACACACACACAACTTTTAACACACAAATCAACACACT 25758

QY 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlyIuTyrAsnMetSerLeu 120

Db 25757 CGTGTGTTGGTTGGCGAGCATACCATAGGCTGGTACTCTGAGTAAATATATGCACTG 25698

QY 121 GlyIuArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140

Db 25697 GGGGAACGCCGTCGCGGTGCGTACCACTATTGCTGGTAAAGCATTAATCAAGCC 25638

QY 141 SerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGlu 160

Db 25637 AGCGTTGAGATTATCAGTTTGGTGAAAGACGCCCTATCGCATTTGGCACAAATGAAGAA 25578

QY 161 AlaTSPserGlnAsnArgAlaGluLeuSerTyr 172

Db 25577 GCATGTCACAAATCGTCGTGCAACTGCTTAT 25542

RESULT 3

AA240352
 ID AA240352 standard; DNA: 519 BP.

AC AA240352;

DT 01-MAR-2000 (first entry)

DE M. catarrhalis BASB019 protein coding sequence #2.

XX BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
 KW genetic mutation screening; antibody production; vaccine; otitis media;
 KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; delayed speech learning; bacteria adhesion prevention;
 KW upper respiratory tract infection; middle ear infection; therapy; ss.

OS Moraxella catarrhalis.

PN WO9957277-A2.

PD 11-NOV-1999.

PF 03-MAY-1999; 99WO-EP03038.

PR 06-MAY-1998; 98GB-0009683.

PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Ruelle J;

DR WPI; 2000-062148/05.

PK P-PSDB; AAY55090.

PT Novel BASB019 polynucleotides and polypeptides from Moraxella

PS catarrhalis used to prepare vaccines against bacterial infections -

XX Claim 10; Fig 2; 101pp; English.

XX This sequence encodes a Moraxella catarrhalis BASB019 protein of the
 CC invention. The sequences can be used for diagnosis of disease, staging of
 CC disease, or determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and for
 CC screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies, and as a


```

QY 1 MetMetLeuHisIleGlnIleAlaIleAlaIleAlaLeuSerValIleuThrPheMet 20
DB 1 ATGATGTTACATATTCAAATTCGCCGCGTCCGCTTATTCGTAACCTTTATG 60
QY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
DB 61 ACAGGCTGTGCAATTAATCAACAGCAAGTATGGTTCCTCTAATGACCCACAGGT 120
QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60
DB 121 TACACTGGGTTATCTACTGCTGTGACCTTTGGTACATATATGATGACTATCAAG 180
QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80
DB 181 ACTTTGGCAAGCAGCGTACCAGTTGGTTATTTGACTTGTATTCATGATTAATA 240
QY 81 ProGlnAlaIleAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
DB 241 CCGCAAGCTCTGCTGCACTTATGACGAACAAGCAACATTTTAAACCAATCAACAGCT 300
QY 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
DB 301 CGATGTTTGGTGCAGGTATACCGATGACCGGTGCTGATGATATATATATGTCAGT 360
QY 121 GlyGluArgArgAlaValAlaValAlaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
DB 361 GGGGAACGCCGTGCGGTGGGTACCACTATTTGCTTGGTAAAGCATTTATCAAGCC 420
QY 141 SerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu 160
DB 421 AGCGTTGAGATTATCACTTTTGGTGAAGAAGCCCTATGCAATTTGGCACAATGAAGA 480
QY 161 AlaTyrSerGlnAsnArgArgAlaGluLeuSerTyr 172
DB 481 GCATGCTCACAAATATCGTCTGCTGAACGTCTTAT 516
RESULT 5
AAZ40354
ID AAZ40354 standard; DNA; 519 BP.
XX
AC AAZ40354;
XX
DT 01-MAR-2000 (first entry)
XX
DE M. catarrhalis BASB019 protein coding sequence #4.
XX
KW BASB019 protein; diagnosis; infectious organism; auditive nerve damage;
KW genetic mutation screening; antibody production; vaccine; otitis media;
KW bacterial infection; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; delayed speech learning; bacteria adhesion prevention;
KW upper respiratory tract infection; middle ear infection; therapy; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO9957277-A2.
XX
PD 11-NOV-1999.
XX
PF 03-MAY-1999; 99WO-EP03038.
XX
PR 06-MAY-1998; 98GB-0009683.
XX
PA (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI; 2000-062148/05.
DR P-PSDB; AAY55092.
XX
PT Novel BASB019 polynucleotides and polypeptides from Moraxella
XX catarrhalis used to prepare vaccines against bacterial infections
PS Claim 10; Fig 2; 101pp; English.

```

```

XX This sequence encodes a Moraxella catarrhalis BASB019 protein of the
CC invention. The sequences can be used for diagnosis of disease, staging of
CC disease, or determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and for
CC screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB019 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies, and as a
CC target for the screening of antimicrobial drugs. The polypeptides can
CC also be used in vaccine formulations, and to identify agonists and
CC antagonists. The polypeptides, antibodies, agonists and antagonists
CC (which are bacteriostatic) are used for the treatment and prevention of
CC diseases including bacterial infection, otitis media in infants and
CC children, pneumonia in the elderly, sinusitis, nosocomial infections and
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infection of the upper respiratory tract and middle ear
CC infection. They are also used in the prevention of adhesion of bacteria
CC to eukaryotic matrix proteins on in-dwelling devices, or to extracellular
CC proteins on wounds, and to thus prevent tissue damage and/or block the
CC normal progression of pathogenesis in infections initiated other than by
CC the implantation of in-dwelling devices or by other surgical techniques.
CC The frequency of Moraxella catarrhalis infections has risen dramatically,
CC and it is no longer common to isolate M. catarrhalis strains that are
CC resistant to standard antibiotics. The BASB019 products of the invention
CC can be used screen for new antibacterial compounds that may target these
CC resistant bacteria.
XX

```

Sequence 519 BP; 137 A; 111 C; 122 G; 149 T; 0 other;

Alignment Scores:

```

Pred. No.: 9,06e-94 Length: 519
Score: 844.00 Matches: 169
Percent Similarity: 98.26% Conservative: 0
Best Local Similarity: 98.26% Mismatches: 3
Query Match: 98.03% Indels: 0
DB: 21 Gaps: 0

```

US-09-674-779B-2 (1-172) x AAZ40354 (1-519)

```

QY 1 MetMetLeuHisIleGlnIleAlaIleAlaIleAlaLeuSerValIleuThrPheMet 20
DB 1 ATGATGTTACATATTCAAATTCGCCGCGTCCGCTTATTCGTAACCTTTATG 60
QY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
DB 61 ACAGGCTGTGCAATTAATCAACAGCAAGTATGGTTCCTCTAATGACCCACAGGT 120
QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLys 60
DB 121 TACGCTGGCTTATCTACACGCTGCTGTGACCTTTGGTACATATATGATGACCGTCAAG 180
QY 61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLys 80
DB 181 GCTTTGGCAAGCAGCGTACCAGTTGGTTATTTGACTTGTATTCATGATTAATA 240
QY 81 ProGlnAlaIleAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAla 100
DB 241 CCGCAAGCTCTGCTGCACTTATGACGAACAAGCAACATTTTAAACCAATCAACAGCT 300
QY 101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
DB 301 CGATGTTTGGTGCAGGTATACCGATGACCGGTGCTGATGATATATATGTCAGT 360
QY 121 GlyGluArgArgAlaValAlaValAlaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAla 140
DB 361 GGGGAACGCCGTGCGGTGGGTACCACTATTTGCTTGGTAAAGCATTTATCAAGCC 420
QY 141 SerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu 160
DB 421 AGCGTTGAGATTATCACTTTTGGTGAAGAAGCCCTATGCAATTTGGCACAATGAAGA 480

```

OY 161 AlaTrpSerGlnAspArgArgAlaGluLeuSerTyr 172
 Db 481 GCATGGTCACAAATCGTCGTGCTGAACGTCTTAT 516

RESULT 6

ID AA006089 standard; DNA; 866 BP.

AC AA006089;

24-JAN-1991 (first entry)

Sequence encoding 16.6kd outer membrane protein (OMP) of H.influenzae.

HI vaccine; OMP; ds.

Haemophilus influenzae.

Key Location/Qualifiers

FT CDS 67..525

FT /*tag= a

EP389925-A.

03-OCT-1990.

20-MAR-1990; 90EP-0105205.

29-MAR-1989; 89US-0330229.

(DYNE-) STATE UNIV NEW YORK.

Murphy TF, Apicella MA;

WPI; 1990-298924/40.

P-PSDB; AAR07145.

Purification of outer membrane protein of haemophilus influenzae

by separation from an insoluble fraction using

detergent-contg., then detergent-free buffers.

Disclosure; Page 8; 22pp; English.

Method claimed produces large quantities of the purified OMP, useful

in raising antibodies for detection, and as a vaccine against

H.influenzae.

Sequence 866 BP; 252 A; 168 C; 175 G; 271 T; 0 other;

US-09-674-779B-2 (1-172) x AA006089 (1-866)

OY 11 AlaAlaLeuSerValLeu-----ThrPheMetThrGlyCysAlaAsn 25

Db 13 TCCAGCTTGTCTCCACTTAACCTAATAAATAAAACATTCATTAGAGAAATTAATGAC 72

OY 26 LysSerThrSerGlnValMetValAlaProAsnAlaProThr----- 39

Db 73 AAATTGTTAAATCATTAATTAGTTGACAGTTCTTACTGACATTAGCGGCTGTGATTC 132

OY 40 -----GlyTyrThrGlyVal 44

Db 133 TCTAACACGATGCTGCAGCAATGTCCTCAAGATTGTCGCGATTACTCT----- 186

OY 45 IleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValValAlaLeuAlaSer 64

||||| |||

Db 187 -----GTTGCTGATCTTCACACAGTTTACACAC----- 216

OY 65 LysLeuProSerLeuValTyrPheAspPheAspSerAspGluLeuLysProGlnAlaAla 84

Db 217 -----GTAATATTGTTGTTGATTAATAATACACATCAACCGGTGAATACGT 261

OY 85 AlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuVal 104

Db 262 CAATCTTAGATGCGCAGCGCATATTTAATGCAACGCGAGCTGCTGAATATTAAGTA 321

OY 105 AlaGlyHisThrAspGluArgGlySerArgGluTyrTrsMetSerLeuGlyGluArg 124

Db 322 GAAGTAATCTGATGAAACGCTGACACCAAGATTAATGATGCTGCTGAATATTAAGTA 381

OY 125 AlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIle 144

Db 382 GCAGATGCACTTAAGGTATTATTGACAGGTAAAGCTTGTGCTGAATATTAAGCACA 441

OY 145 IleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaIlePseGln 164

Db 442 GTATCTTACGCTGAAGAAACCTGACAGTATTAGTACAGATGAAGCTGATATTCTAAA 501

OY 165 AsnArgArgAlaGluLeuSerTyr 172

Db 502 AACGCTGTCGATGTTAGCGTAC 525

RESULT 7

ID AA003869

AA003869 standard; DNA; 737 BP.

AA003869;

30-AUG-1990 (first entry)

H.influenzae DNA fragment containing the pBOMP-1 gene.

outer membrane proteins; pBOMP-1; active immunisation;

typable H influenzae; non-typable H influenzae; ss.

Haemophilus influenzae.

WO9002557-A.

22-MAR-1990.

31-AUG-1989; 89WO-US03779.

01-SEP-1988; 88US-0239572.

21-AUG-1989; 89US-0396572.

(PRAX-) PRAXIS BIOLOGICS IN.

Anilions A, Seld RC, Delch RA, Zlotnick GW, Green BA;

WPI; 1990-115815/15.

P-PSDB; AAR05797.

Outer membrane protein epitopes of Haemophilus influenzae- used in

the prodn. of antibodies, in vaccines and for prodn. of reagents for

diagnosis.

Disclosure; Fig 10; 164pp; English.

See also AA003870, R03948 and AAR03949.

Sequence 737 BP; 225 A; 134 C; 154 G; 224 T; 0 other;

Alignment Scores:

Pred. No.: 1.07e-22 Length: 737
 Score: 264.00 Matches: 51
 Percent Similarity: 68.93% Conservative: 20
 Best Local Similarity: 49.51% Mismatches: 32
 Query Match: 30.66% Indels: 0
 DB: 11 Gaps: 0

US-09-674-779B-2 (1-172) x AAQ03869 (1-737)

QY 70 VALTYRPhEaSPHeaSPSeRasPgiuIlleYsProGlnAlaAlaAlaIleuaspGlu 89
 |||||
 DB 391 GTTATTTCGCTGTTTGAATAAATATGACATTACGCTGGAATACGTTCAATCTTAGATCCG 450
 QY 90 GlnAlaGlnPheLeuThrThraSngInThraSngValleuValAlaGlyHisThraSP 109
 |||||
 DB 451 CACGCTGCATATTAAATGCAACACACACCTGCTAAGATTATGTAAGAGTAACACGAT 510
 QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 |||||
 DB 511 GAACGTGTCACACCAACATACACATCCATTAGCCCAACGCTGCGACATGACGTTAA 570
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 |||||
 DB 571 GGTATTATGCTGCTGAAAGGTGTTGATGCTGTAATTAAGCCACAGATCTTACGGTGAA 630
 QY 150 GluArgProIleAlaPheGlyThrAsnGlnGluAlaTyrSerGlnAsnArgAlaGlu 169
 |||||
 DB 631 GAAAAACCTGCAGATTATGTCATGATGAGAACGTCATATCTTAATAAACCGTCGTCAGTG 690
 QY 170 LeuSerTyr 172
 |||||
 DB 691 TTAGCGTAC 699

RESULT 8
 ID AA81194 standard; DNA; 867 BP.
 AC AA81194;
 XX
 DT 10-OCT-1990 (first entry)
 XX
 DE 16600 dalton outer membrane protein of non-typable H influenzae.
 XX
 KW Haemophilus influenzae; 16600 dalton outer membrane protein (OMP);
 KW pneumonia; meningitis; ss
 XX
 OS Haemophilus influenzae.
 XX
 FH Key Location/Qualifiers
 FT CDS 68..530
 FT /*tag= a
 FT /product=16600 dalton OMP
 XX
 PN EP281673-A.
 XX
 PD 14-SEP-1988.
 XX
 PF 12-NOV-1987; 87EP-0116699.
 XX
 PR 08-OCT-1987; 87US-0092948.
 PR 08-NOV-1986; 86US-0932872.
 XX
 PA (UYN-) RES UNIV NEW YORK.
 XX
 PI Murphy TF, Apicella MA;
 XX
 DR WPI: 1988-258472/37.
 DR P-PSDB; AAP80593.
 XX
 PT plasmids, E.coli, hybridomas and antibodies from H. influenzae protein -
 PT for use as vaccine and detection of Haemophilus influenzae
 XX
 PS Disclosure; ; pp; English.

XX An epitope of the protein encoded by this sequence is recognised by
 CC the 7F3 antibody in a number of isolates. Monoclonal antibodies to
 CC the 16600 D OMP are produced by hybridomas and used to detect
 CC H. influenzae. A suspect sample is contacted with the mAb in the
 CC presence of an indicator such as radiolabelled probe.
 XX

SO Sequence 867 BP; 253 A; 169 C; 174 G; 271 T; 0 other;

Alignment Scores:

Pred. No.: 1.34e-22 Length: 867
 Score: 264.00 Matches: 51
 Percent Similarity: 68.93% Conservative: 20
 Best Local Similarity: 49.51% Mismatches: 32
 Query Match: 30.66% Indels: 0
 DB: 9 Gaps: 0

US-09-674-779B-2 (1-172) x AA81194 (1-867)

QY 70 VALTYRPhEaSPHeaSPSeRasPgiuIlleYsProGlnAlaAlaAlaIleuaspGlu 89
 |||||
 DB 218 GTATTATTGCTTTGATTAATATGACATACCGGTGAATACGTTCAATCTTAGATCGC 277
 QY 90 GlnAlaGlnPheLeuThrThraSngInThraSngValleuValAlaGlyHisThraSP 109
 |||||
 DB 278 CACGACGATATTAAATGCAACGCCACGCTGCTAAAGTATTGTAAGAGTAATACGTA 337
 QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 |||||
 DB 338 GAACGTGTCACACCAACATACACATCCATTAGGCAACGCTGCGACATGACGTTAA 397
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGlu 149
 |||||
 DB 398 GGTATTATGCTGCTGAAAGGTGTTGATGCTGTAATTAAGCCACAGATCTTACGGTGAA 457
 QY 150 GluArgProIleAlaPheGlyThrAsnGlnGluAlaTyrSerGlnAsnArgAlaGlu 169
 |||||
 DB 458 GAAAAACCTGCAGATTATGTCATGATGAGAACGTCATATCTTAATAAACCGTCGTCAGTG 517
 QY 170 LeuSerTyr 172
 |||||
 DB 518 TTAGCGTAC 526

RESULT 9

ID AAQ45440 standard; DNA; 1019 BP.
 AC AAQ45440;
 XX
 DT 18-OCT-1994 (first entry)
 XX

DE Fragment encoding outer membrane protein P6 of H. influenzae.
 XX
 KW Outer membrane protein; vaccine; antibody; treatment; detection;
 KW identification; pneumonia; bacteraemia; meningitis;
 KW postpartum sepsis; acute febrile tracheobronchitis; neonatal sepsis;
 KW acute otitis media; Haemophilus influenzae; ds.
 XX

OS Haemophilus influenzae.
 XX
 FH Key Location/Qualifiers
 FT CDS 220..681
 FT /*tag= a
 FT /product= P6 outer membrane protein
 XX

PN US5300632-A.
 XX

PD 05-APR-1994.
 XX

PF 18-NOV-1986; 86US-0932872.
 XX

PR 18-NOV-1986; 86US-0932872.
 PR 08-OCT-1987; 87US-0092948.

RESULT 11
 AAN80226 standard; DNA; 737 BP.
 ID AAN80226
 AC AAN80226;
 XX
 DT 12-JAN-1991 (first entry)
 XX
 DE Sequence of Haemophilus influenzae Praxis Biologics Outer
 DE Membrane Protein-1 (PBOM-1) gene.
 XX
 KW Vaccine; diagnosis; epitope; passive immunisation; ss.
 OS
 XX Haemophilus influenzae.
 XX
 FH Key Location/Qualifiers
 FT CDS 241..702
 FT /*tag= a
 XX
 PA MO804932-A.
 XX
 PN 14-JUL-1988.
 PD
 XX
 PE 23-DEC-1987; 87WO-US03423.
 XX
 PR 11-DEC-1987; 87US-0132073.
 PR 02-MAR-1987; 87US-0020849.
 PR 31-DEC-1986; 86US-0948364.
 XX
 PA (PRAX-) PRAXIS BIOLOGICS IN.
 XX
 PI Deich RA, Zlotnick G, Green B;
 XX
 DR WPI; 1988-205305/29.
 DR P-PSDB; AAP80665.
 XX
 PT Pure peptide related to epitope of Haemophilus influenzae -
 PT used as immunogens in vaccines and for producing antibodies for
 PT passive immunisation and assays
 XX
 PS Disclosure; Fig 10; 129pp; English.
 XX
 CC A pure antigenic peptide or protein related to an epitope of
 CC Haemophilus influenzae is claimed. Also claimed is a recombinant
 CC vector comprising a DNA sequence coding for an antigenic
 CC determinant of an Haemophilus influenzae outer membrane protein,
 CC the transformed cell, a subunit vaccine in a pharmaceutical carrier,
 CC a method of immunising humans and an assay for Haemophilus influenzae.
 XX
 SQ Sequence 737 BP; 225 A; 133 C; 154 G; 225 T; 0 other;
 Alignment Scores:
 Pred. No.: 4,36e-22 Length: 737
 Score: 259.00 Matches: 50
 Percent Similarity: 67.968 Conservative: 20
 Best Local Similarity: 48.548 Mismatches: 33
 Query Match: 30.084 Indels: 0
 DB: Gaps: 0
 US-09-674-779b-2 (1-172) x AAN80226 (1-737)
 QY 70 valtyrphaspheaspsersasptgltlleysproglinalaialaleuaspclu 89
 Db 391 gtttatttggctttgataaataatgacattactggtgatacgttcaaatcttaacggc 450
 QY 90 glnalaglnphelenthrthrasnglntrraaagvalleuvalalaglyhisthrsp 109
 Db 451 cacgctgcattatttaatacaccacacgctgctaaagattagtagaaggtaacactgat 510
 QY 110 gtuargglyserargglutyrasnmetserleuglygluargargalvalaialarg 129
 Db 511 gaacggtgtacacacgaatatacaaacatcgattagcccaacgctgcagatgcactttaa 570

QY 130 asntyrleuleuglylysclyleasnglnalaaservalglutlleleserphelglylu 149
 Db 571 ggtattattagctgctgtaaaagctgctgataattagcacacattacttaacgctgaa 630
 QY 150 gluargprollalealaphelgylthrasnglnuualatpserclnasargvalaclu 169
 Db 631 gaaaaacctgcagattattaggtcattgcatatctgatttcaaaaacccgtgcactg 690
 QY 170 leusertyr 172
 Db 691 ttacgctac 699
 RESULT 12
 AAT67789
 ID AAT67789 standard; DNA; 540 BP.
 XX
 AC AAT67789;
 XX
 DT 29-JUL-1997 (first entry)
 XX
 DE H. pylori outer membrane protein ORF 31262.aa.
 XX
 KW Outer membrane; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding; compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis; ds.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..540
 FT /*tag= a
 XX
 PN WO9640893-A1.
 XX
 PD 19-DEC-1996.
 PD
 XX
 PE 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTRA) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Meligaerd BL;
 XX
 DR WPI; 1997-052306/05.
 DR P-PSDB; AAM20394.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 Claim 1; Page -: 1481pp; English.
 This sequence encodes a H. pylori outer membrane protein.
 The protein may be used in a vaccine to prevent or treat H. pylori
 infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC Note: This DNA sequence is not reproduced in the specification and
 CC has been derived from the related specification, WO9719098.

SQ Sequence 540 BP; 190 A; 88 C; 128 G; 134 T; 0 other;

Alignment Scores:

Pred. No.:	5,83e-17	Length:	540
Score:	215.50	Matches:	57
Percent Similarity:	50.00%	Conservative:	30
Best Local Similarity:	32.76%	Mismatches:	66
Query Match:	25.03%	Indels:	21
		Gaps:	5

US-09-674-779B-2 (1-172) x AAT67789 (1-540)

```

QY 15 Servalleuthrphemet-----ThrglycysAlaAnlys----- 26
    ||||| :|||:|
DB 13 TCTGATTTAGTTCTTGTAGCTTTTATTGTAGTGTGCTGATCAATAAAGAT 72
QY 27 -----SerThrSerGlnValMetValAlaPro-----Asn 36
    ||||| |||
DB 73 AATAAGACTGTGCTGCGATGTGACCTTAAGCGGTTCAGACTGCGCTGTACTACA 132
QY 37 AlaproThrglytyrThrglyValIleTyThrglyValAlaProLeuValAspAsn 56
    |||
DB 133 GAACCACTCCAGAGAAAGAGAGCCCTTAACAGAGCCAGCTCCAGTGTGAAGAAAG 192
QY 57 GluthrValIleValAlaSerLysLeuProSerLeuValTyPheAspPheAspSer 76
    :|||:|
DB 193 CCGGCTATTGMA-----AGCGGACTATCATCCCTTCTATTATTATTTGATTTGACAA 246
QY 77 AspGluIleLysProGlnAlaIleAlaIleLeuAspGluGlnAlaGlnPheLeuThr 96
    ||||| |||
DB 247 TATAGATCAAGAAATCCGATCAAGACACTTATGATGATCGTCAAAAAGCTAAAGAA 306
QY 97 AsnGlnThrAlaValAlaValAlaGlyHisThrAspGluArgLysSerArgLysTyr 116
    ||| :|||:|
DB 307 AACCAAC---ATGCAAGTCTTTTGGAGGCAATCCGATGATTTGGCTTAGGAAATAC 363
QY 117 AsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyTyLeuLeuGlyLysGly 136
    ||| :|||:|
DB 364 AACCAAGCGCTTGGCGTTAAAGAGACTTGGAGCTGAAGAAACGCTTTAGCATTTAAAGGG 423
QY 137 IleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgProIleAlaPheGly 156
    :|||:|
DB 424 GTAGAAAAGATATGATCAAAACATCATGTTTGGCGAAAGCAACCAATATGCTCAAA 483
QY 157 ThrAsnGluGluAlaTyPheSerGlnAsnArgAlaGluLeu 170
    ||| :|||:|
DB 484 AAACCTAGAGATGTTACAGAGAAACAGAGAGTGTGATGTC 525

RESULT 13
AAT77469
ID AAT77469 standard; DNA; 540 BP.
XX
AC AAT77469;
XX
DT 11-AUG-1997 (first entry)
XX
DE H. pylori outer membrane protein ORF 31262.aa.
XX
KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW detection; antisense; inhibition; ds.
XX
OS Helicobacter pylori.
XX
FH key Location/Qualifiers
FT 1..340
FT CDS /*tag= a
XX
PN MO9719098-Al.
XX
PD 29-MAY-1997.
XX

```

```

PF 15-NOV-1996; 96WO-US18542.
XX
PR 17-NOV-1995; 95US-0561469.
XX
PA (ASTRA ) ASTRA AB.
XX
PI Smith DH;
XX
DR WPI: 1997-298052/27.
XX
DR P-PSDB; AAW24651.
XX
PT Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics
XX
PS Claim 1; Page 108; 235pp; English.
XX
CC This sequence encodes an H. pylori outer membrane protein.
CC Helicobacter pylori has been strongly linked to chronic gastritis and
CC duodenal ulcer disease. The nucleic acid sequences of the invention
CC are used to evaluate compounds, especially activators or inhibitors of
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC sequence. The nucleic acid sequences, and corresponding proteins, are
CC also useful for generating vaccines for immunising subjects against H.
CC pylori or for use in detecting the presence of Helicobacter species in
CC a sample. Antisense nucleic acid sequences of these sequences are
CC used to inhibit expression of a gene from Helicobacter species. H.
CC pylori whole genomic DNA was isolated and nebulised to a median size of
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
CC complementary to the BstXI-cut pmx vectors, while the overhang is not
CC self-complementary. Therefore the linkers will not concatamerise nor
CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
CC were ligated to each of the 20 pmx vectors to construct a series of
CC shotgun subclone libraries. The purified DNA samples were then
CC sequenced.
CC Note: The ORF/protein reference number for this sequence was obtained
CC from the related specification. WO9640893.
XX
SQ Sequence 540 BP; 190 A; 88 C; 128 G; 134 T; 0 other;

Alignment Scores:
Pred. No.: 5.83e-17 Length: 540
Score: 215.50 Matches: 57
Percent Similarity: 50.00% Conservative: 30
Best Local Similarity: 32.76% Mismatches: 66
Query Match: 25.03% Indels: 21
DB: 18 Gaps: 5

US-09-674-779B-2 (1-172) x AAT77469 (1-540)
QY 15 Servalleuthrphemet-----ThrglycysAlaAnlys----- 26
    ||||| :|||:|
DB 13 TCTGATTTAGTTCTTGTAGCTTTTATTGTAGTGTGCTGATCAATAAAGAT 72
QY 27 -----SerThrSerGlnValMetValAlaPro-----Asn 36
    ||||| |||
DB 73 AATAAGACTGTGCTGCGATGTGACCTTAAGCGGTTCAGACTGCGCTGTACTACA 132
QY 37 AlaproThrglytyrThrglyValIleTyThrglyValAlaProLeuValAspAsn 56
    |||
DB 133 GAACCACTCCAGAGAAAGAGAGCCCTTAACAGAGCCAGCTCCAGTGTGAAGAAAG 192
QY 57 GluthrValIleValAlaSerLysLeuProSerLeuValTyPheAspPheAspSer 76
    :|||:|
DB 193 CCGGCTATTGMA-----AGCGGACTATCATCCCTTCTATTATTATTTGATTTGACAA 246
QY 77 AspGluIleLysProGlnAlaIleAlaIleLeuAspGluGlnAlaGlnPheLeuThr 96
    ||||| |||
DB 247 TATAGATCAAGAAATCCGATCAAGAGACTTATGATGATCGTCAAAAAGCTAAAGAA 306
QY 97 AsnGlnThrAlaValAlaValAlaGlyHisThrAspGluArgLysSerArgLysTyr 116
    ||| :|||:|
DB 307 AACCAAC---ATGCAAGTCTTTTGGAGGCAATCCGATGATGATTTGGCTTAGGAAATAC 363

```

QY 117 AsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGly 136
 DB 364 AACCAAGCGCTGGCTTAAGAGCGCTTGAAGCGGAAACGCTTAGTATTAAGCGG 423
 QY 137 IleAsnGlnAlaSerValGluIleIleSerPheGlyGluArgProIleAlaPheGly 156
 DB 424 GTAGAAAAGATATGATCAAAACCATCATGCTTTGGCGAAACCAACCAATGCGTCOA 483
 QY 157 ThrAsnGluGluAlaTyrSerGlnAsnArgArgAlaGluLeu 170
 DB 484 AAACTAGAGATGTTATACAGAAAACAGAGAGTGAATGTC 525
 RESULT 14
 AAF25593 standard; DNA; 540 BP.
 AC AAF25593;
 DT 05-APR-2001 (first entry)
 XX H. pylori HPS144 encoding DNA.
 DE Microbial infection; antibacterial; Helicobacter pylori infection;
 KW vaccine; screening; ds.
 XX Helicobacter pylori.
 OS WO200073502-A2.
 PN 07-DEC-2000.
 XX 31-MAY-2000; 2000WO-EP05024.
 PE 31-MAY-1999; 99DE-1024965.
 PR 17-JUN-1999; 99DE-1027740.
 PR 21-JUL-1999; 99DE-1034029.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (CREA-) CREATOGEN GMBH.
 FA
 XX
 PI Apfel H, Fuchs TM, Glibbs CP, Hueck CJ, Meyer TF;
 DR WPI: 2001-049948/06.
 DR P-PSDB: AAB46316.
 PT Preparing an agent for diagnosis or control of microbial infection,
 PT useful particularly against Helicobacter, based on identification of
 XX essential genes in defective mutants -
 PS Claim 37; Page 252-253; 366pp; German.
 XX
 CC This invention describes a novel preparation of an agent (A) for
 CC detection, prevention and/or treatment of microbial infection by:
 CC (i) identifying essential genes (I) and corresponding polypeptides
 CC (II); (ii) identifying compounds that are directed against (II) and
 CC inactivate the microbe; (iii) testing these for suitability for use; and
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises
 CC preparation of gene-deficient microorganisms by conditional antisense
 CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM).
 CC then determining viability and/or survival of the deficient organisms.
 CC The products of the invention have antibacterial activity. (A) which may
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their
 CC treatment or prevention of infection by Helicobacter pylori. Particularly
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method
 CC identifies essential genes, including those that have homologs in other
 CC species, so identified (A) should have a broad spectrum of activity. Many
 CC gene-deficient cells can be screened quickly, in an automated process,
 CC and the identified genes can be used for screening without purification.
 XX
 XX Sequence 540 BP; 192 A; 89 C; 125 G; 134 T; 0 other;

Alignment Scores:
 Pred. No.: 5.83e-17 Length: 540
 Score: 215.50 Matches: 56
 Percent Similarity: 51.72% Conservative: 34
 Best Local Similarity: 32.18% Mismatches: 63
 Query Match: 25.03% Indels: 21
 DB: 22 Gaps: 5
 US-09-674-779b-2 (1-172) x AAF25593 (1-540)
 QY 15 SerValLeuThrPheMet-----ThrGlyCysAlaAspLysSerThr 28
 DB 13 TCTGTATTAGTTAGTTCTTGGTAGCTTTTATTGTAAGCTGCTGATGATATAATGAT 72
 QY 29 SerGlnValMetValAlaPro--AsnAlaProThrGlyTyrThrGlyValIleTyrThr 47
 DB 73 AATAAGACTGTGGCGCGCATGTGAGTGTCTAAACGCTTCAAGCTGACCTGTTACTACA 132
 QY 48 -----GlyValAlaProLeuValAspAsnAsp 56
 DB 133 GAACCAAGCTCCAGAGAAAGAGCCTTAAACAGACCAGCTCCAGTGTGAGAAAGAAA 192
 QY 57 GluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSer 76
 DB 193 CCGGCTGTGAG-----AGCGGACTATCATCGCTTCATTATTATTGATTTCACAAG 246
 QY 77 AspGluIleLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThr 96
 DB 247 TATGAATCAAGAAATCCGATCAAGAGACTTATGATGATGCTGCAAAAGCTAAAGA 306
 QY 97 AsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGluTyr 116
 DB 307 AACCCAC--ATCGAAGTGTCTTTTGAAGGCAATACCGATGATTTGGCTTAGCGCAATAC 363
 QY 117 AsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGly 136
 DB 364 AACCAAGCGCTGGCTTAAGAGCGCTTGAAGCGGAAACGCTTAGTATTAAGCGG 423
 QY 137 IleAsnGlnAlaSerValGluIleIleSerPheGlyGluArgProIleAlaPheGly 156
 DB 424 GTAGAAAAGATATGATCAAAACCATCATGCTTTGGCGAAACCAACCAATGCGCCOA 483
 QY 157 ThrAsnGluGluAlaTyrSerGlnAsnArgArgAlaGluLeu 170
 DB 484 AAACTAGAGATGTTATTAAGAAAACAGAGAGTGAATGTC 525
 RESULT 15
 AAT68048
 ID AAT68048 standard; DNA; 561 BP.
 XX
 AC AAT68048;
 DT 21-JUL-1997 (first entry)
 XX
 DE H. pylori outer membrane protein ORF 07gp31516orf4.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
 XX
 OS Helicobacter pylori.
 XX
 FT Key Location/Qualifiers
 FT CDS 1..561
 FT /*tag= a
 FT /*note= "no stop codon given"
 XX
 XX WO9640893-A1.
 XX 19-DEC-1996.
 PD
 XX
 PF 06-JUN-1996; 96WO-US09122.

```

XX 01-APR-1996; 9605-0630405.
PR 07-JUN-1995; 9505-0487032.
XX
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
XX
XX WPI: 1997-052306/05.
DR P-PSDB; AAM20795.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) -- useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
XX Claim 1; Page 866; 1481pp; English.
XX
CC This sequence encodes a H. pylori outer membrane protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SQ Sequence 561 BP; 196 A; 88 C; 131 G; 146 T; 0 other;

Alignment Scores:
Pred. No.: 6.15e-17 Length: 561
Score: 215.50 Matches: 57
Percent Similarity: 50.00% Conservative: 30
Best Local Similarity: 32.76% Mismatches: 66
Query Match: 25.03% Indels: 21
DB: 18 Gaps: 5

US-09-674-779b-2 (1-172) x AAT68048 (1-561)
QY 15 SerValLeuThrPheMet-----ThrGlyCysAlaAsnLys----- 26
DB 37 TCTGTAATTTAGCTTCTTGAGACCTTTTATGTGTAATGGCTGCTATCAATAAGAT 96
QY 27 -----SerThrSerGlnValMetValAlaPro-----Asn 36
DB 97 AATAAGACTGCTGCTGGAGATGTCAGCATAAGCGCTGTCAGACTGCTGTTACTACA 156
QY 37 AlAProThrGlyTYrThnGlyValILleTYrThnGlyValAlaProLeuValAspAsn 56
DB 157 GAACCACTCCAGAGAAAGAGACCTAAACAGAGAGCCAGCTCAGGGTGAAGAAAG 216
QY 57 GluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTYrPheAspPheAsp 76
DB 217 CCGGCTATTGAA-----ACGGGACATATCATCGCTCTATTATTTGATTTGACAA 270
QY 77 AspGluLeuLysProGlnAlaAlaAlaILleuAspGluGlnAlaGlnPheLeuThr 96
DB 271 TATGAGATCAAGATCCGATCAAGACACTTAAATAGATCCGCAAAAAGCTAAAGAA 330
QY 97 AsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlu 116
DB 331 AACCAAC---ATGCAAGTCTTTTGAAGGCAATACCGATGAAATTGGCTCTAGCGAT 387
QY 117 AsnMetSerLeuGlyArgArgAlaValAlaValArgAsnTYrLeuLeuGlyLysGly 136
DB 388 AACCAAGCGCTTGCGTTAAAGAGACATTGAGCGTGA AAAACGCTTATGATCAATA 447
QY 137 IleAsnGlnAlaSerValGluILleIleSerPheGlyGluArgProILleAlaPheGly 156

```

```

DB 448 GTAGAAAAAGTATGATCAACAAACCATGTTGGCAAGCAACCAAAATGCTCCAA 507
QY 157 ThrAsnGluGlnAlaTrpSerGlnAsnArgArgAlaGluLeu 170
DB 508 AAACAGAGATGTATACAGAGAAACAGCAAGATGGATGTC 549

RESULT 16
AAV90653
ID AAV90653 standard; DNA; 770 BP.
XX
XX AAV90653;
AC
XX
XX 18-FEB-1999 (first entry)
DE
XX Nucleotide sequence of clone Y175.ASM from cluster 2.
KW
XX Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma; ss.
XX
XX Helicobacter pylori.
XX
XX W09849314-A2.
XX
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08487.
XX
XX 14-OCT-1997; 97US-0061958.
XX
XX 25-APR-1997; 97US-0045107.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX
XX Chow TP, Fry KE, Lim MY, McAttee CP;
XX
XX WPI: 1999-009433/01.
XX
XX New Helicobacter pylori antigens and related nucleic acid sequences
XX - useful in serological diagnosis and protective vaccines, providing
XX long-lasting immune response
XX
XX Claim 27; Page 143-144; 402pp; English.
XX
XX The present sequence encodes a Helicobacter pylori antigenic protein
XX that is characterised by immunoreactivity with H. pylori positive
XX antisera. The proteins are highly immunogenic and induce a long-lasting
XX immune response that persists even after antimicrobial treatment. In
XX antibody-detection assays, on sera, plasma, urine, saliva etc., they are
XX highly sensitive and specific. The specification also describes 69
XX previously unrecognised immunogenic cluster families. H. pylori antigens
XX are used to detect H. pylori-specific antibodies, for diagnosing
XX infection or to confirm eradication of infection, and in vaccines to
XX protect against H. pylori infection and related diseases (gastritis,
XX peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ Sequence 770 BP; 264 A; 132 C; 164 G; 210 T; 0 other;

Alignment Scores:
Pred. No.: 2.95e-16 Length: 770
Score: 211.50 Matches: 54
Percent Similarity: 52.02% Conservative: 36
Best Local Similarity: 31.21% Mismatches: 72
Query Match: 24.56% Indels: 11
DB: 20 Gaps: 4

US-09-674-779b-2 (1-172) x AAV90653 (1-770)
QY 8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThnGlyCysAlaAsnLysSer 27
DB 40 TCTTCTGCAATTAGTCTTGTGATCCTTTTATTTGAGTGGCTGCTGATGATTAATG 99
QY 28 ThrSerGlnValMetValAlaPro---AsnAlaProThnGlyTYrThnGlyValILleTYr 46

```


[illegible]

US-09-674-779B-2 (1-172) x AAT74194 (1-435)

QY 50 AlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerLeu 69
 Db 67 GCTCAGAGGGTTGAGAAAGCCGCGTGTGAG-----AGCGGACTATCATCGCTTC 120
 QY 70 ValTyrPheAspPheAspSerAspGluLeuLysProGlnAlaAlaIleLeuAspGlu 89
 Db 121 ATTATATTTGATTTTGACAGTATGATCAATCAAGATCCGATCAAGAGACTTTAGATGAG 180
 QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlnHisThrAsp 109
 Db 181 ATCTGTGAAAAAGCTAAAGAAACCCAC---ATGCAAGTCTTTGGAGCAATACCGAT 237
 QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyLysArgArgAlaValAlaValArg 129
 Db 238 GAATTTGGCTCTACCGAATTCACCAAGCGCTTGGCTTAAAGGACTTTGACCGTGA 297
 QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluLeuIleSerPheGlyLys 149
 Db 298 AACCTTAACTTATTAAAGGGTGAAGAAAGATATATCAAAACATCACTTTTGGTGA 357
 QY 150 GlnArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArgArgAlaGlu 169
 Db 358 ACCAAACCCAAATGACCCCAAAAACTAGAGAGTGTATTAAGAAAAAGAGAGTGAT 417
 QY 170 Leu 170
 Db 418 GTC 420

RESULT 20
 ID ABL91202 standard; DNA; 579 BP.
 AC ABL91202:
 DT 29-JUL-2002 (first entry)
 XX Chlamydia pneumoniae cp7090 ORF DNA, SEQ ID NO:38.
 DE Chlamydia pneumoniae cp7090 ORF DNA, SEQ ID NO:38.
 XX Chlamydia pneumoniae.
 OS Chlamydia pneumoniae.
 XX Key Location/Qualifiers
 FT CDS 1..579
 FT /tag- a
 FT /product- "cp7090"
 FT sig_peptide 1..60
 FT /tag- b
 FT mat_peptide 61..576
 FT /tag- c
 FT /product- "Mature protein"
 XX
 XX W0200202606-A2.
 PD 10-JAN-2002.
 PF 03-JUL-2001; 2001MO-IB01445.
 PR 03-JUL-2000; 2000GB-0016363.
 PR 11-JUL-2000; 2000GB-0017047.
 PR 21-JUL-2000; 2000GB-0017983.
 PR 07-AUG-2000; 2000GB-0019368.
 PR 18-AUG-2000; 2000GB-0020440.
 PR 14-SEP-2000; 2000GB-0022583.
 PR 10-NOV-2000; 2000GB-0027549.
 PR 22-DEC-2000; 2000GB-0031706.

XX (CHIR-) CHIRON SPA.
 PA Rattl G, Grandi G;
 XX WPI: 2002-154726/20.
 XX N-FSDB; ABB90544.
 DR Novel Chlamydia pneumoniae protein useful in the manufacture of a
 PT medicament for treatment or prevention of infection due to Chlamydia,
 PT preferably Chlamydia pneumoniae, and for diagnostic purposes
 PS Claim 5; Page 60-61; 364pp; English.
 XX Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
 CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
 CC them. The proteins are predicted to be immunogenic and may therefore be
 CC useful in vaccine production and for diagnostic purposes. Chlamydia
 CC pneumoniae is a common cause of respiratory disease in humans, and is
 CC also involved in the development of cardiovascular diseases such as
 CC atherosclerosis, coronary artery disease, carotid artery stenosis,
 CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
 CC claudication and stroke. The proteins and nucleic acids of the invention
 CC may be used in vaccines and pharmaceutical compositions for the
 CC prevention or treatment of chlamydial infections, particularly Chlamydia
 CC pneumoniae infections. The proteins may also be used in the detection of
 CC Chlamydia pneumoniae, and the nucleic acids may be used in the detection of
 CC pneumoniae gene expression. The present sequence represents a
 CC specifically claimed DNA which encodes a Chlamydia pneumoniae protein of
 CC the invention.
 XX

SO Sequence 579 BP; 184 A; 131 C; 109 G; 155 T; 0 other;

Alignment Scores:
 Pred. No.: 1,42e-15 Length: 579
 Score: 204.50 Matches: 46
 Percent Similarity: 58.82% Conservative: 24
 Best Local Similarity: 38.66% Mismatches: 46
 Query Match: 23.75% Indels: 3
 DB: 24 Gaps: 2

US-09-674-779B-2 (1-172) x ABL91202 (1-579)

QY 54 AspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerLeu-----ValTyr 71
 Db 202 GATTCGAAGAGAAAGAAACCAATACAGCAAGCCAGTTCGATATATATACACC 261
 QY 72 PheAspPheAspSerAspGluLeuLys--ProGlnAlaAlaIleLeuAspGluGln 90
 Db 262 TTTCCTACAGACAGCTATACATTAAGTGAAGAACCTTGCTTCCACCACTTG 321
 QY 91 AlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlnHisThrAspGlu 110
 Db 322 GTTCACATCATGAGAAAGAAACCCGAAAGCTACCTGATCATTTGAAGGCATACGACGAG 381
 QY 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyLysArgArgAlaValAlaValArgAsn 130
 Db 382 CGTGGAGCTGCATCTCTATACCTTGCTTTAGACAGCAGCAGCAATGCGATTAAAGAG 441
 QY 131 TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluLeuIleSerPheGlyGluGln 150
 Db 442 CATCTCCGAAGCAGAGGAAATCTTCGAGATGCTGATCTACTATTCTTCGGAAGAAAGAA 501
 QY 151 ArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArgArgAlaGlu 169
 Db 502 CATCTTTAAATTCGGGACACACAGCAAGTACGATGCAACAAATCCCGTACAGAG 558

RESULT 21
 AAD08593
 ID AAD08593 standard; DNA; 779 BP.
 AC AAD08593;

XX 04-SEP-2001 (first entry)
 XX Chlamydia pneumoniae omp P6 precursor gene.
 DE Chlamydia pneumoniae omp P6 precursor gene.
 XX Omp P6 precursor; outer membrane protein; Chlamydia infection;
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; atherosclerosis; asthma;
 KW vaccine; antibody; ds.
 XX Chlamydia pneumoniae.
 OS Chlamydia pneumoniae.
 XX Key Location/Qualifiers
 FT CDS 101..679
 FT /*tag= a
 FT /product= "Omp P6 precursor"
 PN MO200146224-A2.
 XX 28-JUN-2001.
 XX 20-DEC-2000; 2000WO-CA01534.
 XX 22-DEC-1999; 99US-0171525.
 XX (AVET) AVENTIS PASTEUR LTD.
 XX Murdin AD, Oomen RP, Wang J, Dunn P;
 DR WPI; 2001-408630/43.
 DR P-PSDB; AAE04322.
 XX Chlamydia polypeptides and polynucleotides for preventing, diagnosing
 PT and treating Chlamydia infection in mammals, especially humans -
 XX Claim 2; Fig 1; 74pp; English.
 XX The present sequence is Chlamydia pneumoniae outer membrane
 CC protein (omp) P6 precursor gene. The omp P6 precursor polynucleotide,
 CC polypeptide and its antibody are useful for detecting, preventing and
 CC treating Chlamydia infections e.g. pneumonia, upper respiratory tract
 CC diseases (e.g. bronchitis, sinusitis) and acute respiratory diseases
 CC (e.g. cough, sore throat). C. pneumoniae infection is also observed to
 CC be associated with atherosclerosis and asthma. Immunogenic fragments
 CC of omp P6 precursor protein and their corresponding DNAs are useful in
 CC vaccine composition. The polynucleotide is useful in the construction
 CC of attenuated Chlamydia strains that can over express the polynucleotide
 CC or express it in a non-toxic, mutated form. The omp P6 probes are useful
 CC in diagnostic tests as capture or detection probes and the primers are
 CC useful in diagnostic methods involving PCR (polymerase chain reaction).
 CC The antibody is useful for purifying polypeptide.
 XX Sequence 779 BP; 240 A; 175 C; 150 G; 214 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 2,14e-15 Length: 779
 Score: 204.50 Matches: 46
 Percent Similarity: 58.82% Conservative: 24
 Best Local Similarity: 38.66% Mismatches: 46
 Query Match: 23.75% Indels: 3
 DB: 22 Gaps: 2
 US-09-674-779B-2 (1-172) x AAD08593 (1-779)
 OY 54 ASPASNAAPGIUThVallySAlaleuAlaserLysleuProserLeu-----ValTyr 71
 Db 302 GATTCGAAGAGAGAAAAACATACAGTCAGCCAAAGTTCGATTTTCGTAATATACACC 361
 OY 72 PheaspPheaspSeraspGluIleLys---ProGlnAlaAlaAlaIleLeuaspGluGln 90
 Db 362 TTTCGTACAGACAGCTATACATTAAGGTGAAGAGACCTTCGATTCACGAACTTG 421
 OY 91 AlagInPheLeuThThAsnGlnThThAlaArgValLeuValAlaGlyHisThThAspGlu 110

Db 422 GTTCACTACATGAGAAAAACCCGAAGCTACACTGTACATTGAGGCGATACGAGAG 481
 OY 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyLuarArgAlaValAlaValArgAsn 130
 Db 482 CGTGGAGCTGCATCTTATACCTTGTAGACACGACGACCAATTCGATTAAAGAG 541
 OY 131 TyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGln 150
 Db 542 CATCTCCGAAGACAGAGAAATCTCGAGATGCTATCTACTACTATTCTCAGGAAAGAA 601
 OY 151 ArgProIleAlaPheGlyThAsnGluLuarIlePheSerGlnAsnArgAlaGlu 169
 Db 602 CATCCTTTAAATTCGCGACACAGCACTGCAATGCAAAATGCCGTACAGAG 658
 RESULT 22
 ID AAX91990 standard; DNA; 1230025 BP.
 AC AAX91990;
 XX 13-SEP-1999 (first entry)
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.
 XX Chlamydia pneumoniae.
 OS Chlamydia pneumoniae.
 XX WO9927105-A2.
 XX 03-JUN-1999.
 XX 20-NOV-1998; 98WO-IB01890.
 XX 04-NOV-1998; 98US-0107078.
 XX 21-NOV-1997; 97FR-0014673.
 XX (GEST) GENSET.
 XX Griffais R;
 DR WPI; 1999-357842/30.
 XX Genome sequence of Chlamydia pneumoniae
 PT Claim 1; Page 291-611; 1912pp; English.
 XX The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
 SQ
 Alignment Scores:
 Pred. No.: 6.07e-11 Length: 1230025
 Score: 204.50 Matches: 46
 Percent Similarity: 58.82% Conservative: 24
 Best Local Similarity: 38.66% Mismatches: 46
 Query Match: 23.75% Indels: 3
 DB: 20 Gaps: 2
 US-09-674-779B-2 (1-172) x AAX91990 (1-1230025)

```

OY 54 AspsnaspGIUTThrValysAlaLeuAlaSerLysLeuProSerLeu-----ValTyr 71
Db 889697 GATTCACAAAGAGAAAAACATACAGTCAAGCCAGATTTGCAATATCACC 889638
OY 72 PheaspPheaspSerAspGluIleLys---ProGlnAlaAlaIleLeuAspGlu 90
Db 889637 TTTGCTACAGACAGTATACATTAAGAGAACCTTGGATTCACAGAACTTG 889578
OY 91 AlaGlnPheLeuThrThrasnGlnThrAlaArgValLeuAlaGlyHisThrAspGlu 110
Db 889577 GTTCACATACATGAGAAAAAACCAGAACCTACCTACGTACATGTAAGGCACTGACGAG 889518
OY 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyLuarArgAlaValAlaArgAsn 130
Db 889517 CGTGAGACTGCATCCTTAATACCTTGTAGAGACGACGAGCCATGCGATTAAAGAG 889458
OY 131 TyrLeuLeuGlyLysGlyIleasnGlnAlaSerValIleLeuSerPheGlyGlu 150
Db 889457 CATCTCGAAGACGAGGAAATCTGACAGATGCTTATCTACTATTCTCAGGAAAAAGAA 889398
OY 151 ArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArgAlaGlu 169
Db 889397 CATCTTAAATTCGGGACACACAGACTAGCATGCAACAAATCGCCCTACAGAG 889341
RESULT 23
ID AA201425 standard; DNA; 1038602 BP.
AC AA201425;
XX 07-OCT-1999 (first entry)
XX Complete genome sequence of Chlamydia trachomatis.
DE
XX
XX Vaccine; eye disease; conjunctival trachoma; nongonococcal trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perlepatitis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX Chlamydia trachomatis.
OS
XX MO9928475-A2.
XX 10-JUN-1999.
XX
XX 27-NOV-1998; 98WO-IB01939.
XX
XX 04-NOV-1998; 98US-0107077.
XX 28-NOV-1997; 97FR-0015041.
XX 17-DEC-1997; 97FR-0016034.
XX
XX (GEST ) GENSET.
XX
XX Griflais R;
XX
XX WPI; 1999-371125/31.
XX
XX Genome sequence of Chlamydia trachomatis
XX
XX Claim 1; Page 373-656; 1755pp; English.
XX
XX The present sequence represents the complete genome of Chlamydia
XX trachomatis. Open reading frames (ORFs) of the genome encode
XX polypeptides AY36754-Y37949. The polypeptides can be used as vaccines
XX against Chlamydia trachomatis. Antisense and ribozyme sequences can also
XX be used to control growth of the microorganism. Chlamydia trachomatis is
XX responsible for a large number of diseases, e.g. eye diseases such as
XX conjunctivitis; genital diseases such as nongonococcal urethritis,
XX epididymitis, cervicitis, salpingitis, perlepatitis, Bartholinitis;
XX pneumopathy in breast feeding infants; and venereal
XX lymphogranulomatosis. The polypeptides of the invention may be of use in
XX treating these diseases.

```

```

XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;
Alignment Scores:
Pred. No.: 4.55e-10 Length: 1038602
Score: 196.50 Matches: 49
Percent Similarity: 48.928 Conservative: 19
Best Local Similarity: 35.258 Mismatches: 54
Query Match: 22.828 Indels: 17
DB: Gaps: 2
US-09-674-779B-2 (1-172) x AA201425 (1-1038602)
OY 48 GlyValAlaProLeuValAspsnaspGIUTThrValysAlaLeuAlaSerLysLeuPro 67
Db 833233 GGATTCGATCCTTTCTACTCCATGAGAAATTCAAACAGCTTTGTGGAATTTGAT 833174
OY 68 SerLeu-----ValTyr 71
Db 833173 TCCAAAGAGACAGCTGTACAAAACGACGACAGAGTACCTTTCCGAATATTCAC 833114
OY 72 PheaspPheaspSerAspGluIleLysProGln---AlaAlaAlaIleLeuAspGlu 90
Db 833113 TTGCTACAGATAGTATCTTAAAGAGAGATACCTCAGCATTTCTGCAAGCTTA 833054
OY 91 AlaGlnPheLeuThrThrasnGlnThrAlaArgValLeuAlaGlyHisThrAspGlu 110
Db 833053 GTTCGTCATTCATTAATCTCTTAAGCTTACGCTATATATAGAGGCCATACAGATGAA 832994
OY 111 ArgGlySerArgGluTyrAsnMetSerLeuGlyLuarArgAlaValAlaArgAsn 130
Db 832993 CGTGAGCTGCGACGCTTATACCTTACCTTACGAGCTGCTGCGAATGCTGTAACAA 832934
OY 131 TyrLeuLeuGlyLysGlyIleasnGlnAlaSerValIleLeuSerPheGlyGlu 150
Db 832933 TACCTCAACAAACAGAAATCGTCGACAGCGCTTATTCACATTTCTTACGAAAAAGAA 832874
OY 151 ArgProIleAlaPheGlyThrAsnGlnGluAlaTrpSerGlnAsnArgAlaGlu 169
Db 832873 CATCTGTCATCCAGGCCATATGAAATTAATTCCTGGCAACAAATCGTCTACTGAA 832817
RESULT 24
AAF30043
ID AAF30043 standard; DNA; 675 BP.
AC AAF30043;
XX
XX 23-APR-2001 (first entry)
XX
XX Moraxella catarrhalis DNA encoding BASB13 protein.
XX
XX BASB13; infection; otitis media; pneumonia; gene therapy;
KW diagnosis; antibacterial; antimicrobial; genetic immunisation;
KW vaccine; ss.
XX
XX Moraxella catarrhalis.
XX
XX Key Location/Qualifiers
XX mat_peptide 1..672
XX
XX /*tag= a
XX /note= "a polynucleotide having this sequence is
XX also specifically claimed in Claim 10"
XX
XX MO200100836-A1.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-EP05851.
XX
XX 25-JUN-1999; 99GB-0015044.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```


PA (BEHM) BEHRINGER AG.
 XX Broecker M, Dondley H, Hungerer K, Knapp B, Ulrich B.
 PI Von Specht BU;
 XX WPI: 1996-279559/29.
 DR P-PSDB; AAR99626.
 XX
 PT Fusion proteins for Pseudomonas aeruginosa vaccines - contg.
 PT fragments of outer membrane proteins I and F.
 XX
 PS Disclosure: Page 13-14; 23pp; English.
 CC A CDNA sequence (AAT32600) codes for a fusion (AAR99626) between amino
 CC acids 192-342 of the outer membrane protein F (OprF) and amino acids
 CC 21-83 of outer membrane protein I (OprI) of Pseudomonas aeruginosa
 CC ATCC 33354. A glutathione-S-transferase fusion with the hybrid gene
 CC was expressed in Escherichia coli and Saccharomyces cerevisiae
 CC transformants. The oprF-oprI hybrid protein, and antibodies raised
 CC against it, conferred protection against P. aeruginosa infection in
 CC laboratory animals. The hybrid protein was significantly more
 CC immunogenic than an oprI-oprF fusion (AAR99627).
 CC
 XX Sequence 645 BP; 148 A; 210 C; 183 G; 104 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 5.76e-12 Length: 645
 Score: 175.50 Matches: 49
 Percent Similarity: 46.58% Conservative: 19
 Best Local Similarity: 33.56% Mismatches: 68
 Query Match: 20.38% Indels: 11
 DB: 17 Gaps: 2
 US-09-674-779b-2 (1-172) x AAT32600 (1-645)
 QY 37 AlaProThrglyTYrThrglyValIleTyThrglyValAlaProLeuValAlaAspAsnAsp 56
 Db 26 GCTCCGACCTCGACACAGCAGCGGCTGCGACAAAC-GTCGACAAAGTCCCGGACCCCG 84
 QY 57 GluThrValAlaLeuAlaSerLysLeuProSerLeu----- 69
 Db 85 GCCAACGTCACCGTTGAGCGCCGCTGCTGCCGAGTCTGACCGCTACAG 144
 QY 70 -----ValTYrPheAspPheAspSerAspLysLeuProLysLeuValAlaAlaLeu 87
 Db 145 CTGACGCGAAGTTCGACTTCGACCAAGTCAAGTCAAGAGAAAGAACGCTACGCTGACATC 204
 QY 88 AspGluGlnAlaGlnPheLeuThrThraSnglInThraAlaArgValLeuValAlaGlyHis 107
 Db 205 AAGAACCTGGCCGACTTCATGACAGCAGTACCCGCTCCACCTTCACACCGCTTGAAGTCAT 264
 QY 108 ThrAspGluArgGlySerArgGluTYrAsnMetSerLeuGluArgAlaValAla 127
 Db 265 ACCGACCTCCGCTGACCGACGCTTACACCAAGAGAGCTGTCGACGCTGCCAACGCC 324
 QY 128 ValArgAsnTYrLeuLeuGlyLys--GlyTLeaSnglAlaSerValGluIleIleSer 146
 Db 325 GTTCGTACAGTACAGTCAAGTCAAGTGTGAGTGTGCGTGAACGCTGTGGT 384
 QY 147 PheGluGluArgProIleAlaPheGlyThraSnglGluAlaTYrPserGlnAsnArg 166
 Db 385 TACGCCGAGTCCCGCGCTGCGACCAACGCCACCGCTGAAAGCGCGCTATCAACCGT 444
 QY 167 ArgAlaGluLeuSerTYr 172
 Db 445 CGCGTTGAAGCAGCCAC 462
 RESULT 28
 AA054144
 ID AA054144 standard; DNA; 4274 BP.
 XX
 AC AA054144;
 XX

DT 24-JUN-1994 (first entry)
 XX
 DE Sequence of plasmid pRW3.
 XX
 KW Outer membrane protein; OprF; plasmid pRW3; ss.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN W09324636-A.
 PD 09-DEC-1993.
 PF 27-MAY-1993; 93WO-CA00227.
 PR 29-MAY-1992; 92US-0891495.
 PA (UWBR-) UNIV BRITISH COLUMBIA.
 PI Hancock RM, Wong R;
 XX WPI: 1993-405827/50.
 DR
 XX
 PT Use of OprF protein - in expression of heterologous
 PT oligopeptide(s) on gram-negative bacterial cell surface to
 PT produce live vaccines and to map antigenic epitope(s)
 XX
 PS Example: Fig 2A; 45pp; English.
 CC Plasmid pRW3 contains the whole P.aeruginosa outer membrane protein
 CC OprF gene with a mutated promoter in pT719R. It was isolated for
 CC the creation of OprF epitope insertion vectors which include a
 CC promoter and a DNA sequence encoding at least the amino terminal
 CC portion of a P.aeruginosa OprF. Inserted in the DNA sequence are
 CC one or more unique restriction sites for insertion of one or more
 CC DNA sequences encoding a protein of interest.
 CC
 XX Sequence 4274 BP; 1048 A; 1135 C; 1135 G; 956 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 2.14e-10 Length: 4274
 Score: 172.00 Matches: 54
 Percent Similarity: 45.09% Conservative: 24
 Best Local Similarity: 31.21% Mismatches: 79
 Query Match: 19.98% Indels: 17
 DB: 14 Gaps: 3
 US-09-674-779b-2 (1-172) x AA054144 (1-4274)
 QY 7 IleAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrglyCysAlaAsnLys 26
 Db 3252 CTGGCCTGGGCTGCGCTTCACTTCGTTGTCGAAAGCCGCTCCGACCGG 3311
 QY 27 SerThrSerGlnValMetValAlaProAsnAlaProThrglyTYrThrglyValIleTy 46
 Db 3312 -----TTGCCGACGTTTGTCTCGACCTCGACAAACGACGCGCTTTCG 3353
 QY 47 ThrglyValAlaProLeuValAlaAspAsnAspGluThrValLysAlaLeuAlaSerLysLeu 66
 Db 3354 ACMAAC-GTCGACAAAGTCCCGGATACCCCGGCAACGTCACCTGACGCCAACGCTGC 3412
 QY 67 ProSerLeu-----ValTYrPheAspPheAspSerAsp 77
 Db 3413 CCGGCTGTGCGCGAGACTCTACGCTGACGCTGACGCTGAATTCGACTGCAAGTCC 3472
 QY 78 GluIleLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThraS 97
 Db 3473 AAGGTCAAGAGAACGACGCTACGCTGACACAACTGCTGACTTCATGAAAGCAGTAC 3532
 QY 98 GlnThraAlaArgValLeuValAlaGlyHisThraSnglArgGlySerArgGluTYrAsn 117
 Db 3533 CCGTCCACTTCCACGACCGCTTGAAGTCAACGACGCTCGTGGCAGCAGCGCTTACAC 3592
 QY 118 MetSerLeuGlyGluArgArgAlaValAlaValArgAsnTYrLeuLeuGlyLys--Gly 136

Db 3593 CAGAGCTGTCGAGCGCTGCGCCACGCGCTGTCGACGACTGTCACGACGACGCT 3652
 QY 137 ILeaNgInAlaSerValGluIleIleSerPheGlyGluGuaArgProIleAlaPheGly 156
 Db 3653 GTAAGAGTGTCCCGTGAACGCTGTGTGCGGACGACGCGCGGTTGCCGACMAC 3712
 QY 157 ThrAsnGluAlaTrpSerGlnAsnArgArgAlaGlu 169
 Db 3713 GCCACCGCTGAAGCCGCGGTATCAACCGTCGCTTAA 3751

RESULT 29

AAT32599
 ID AAT32599 standard; cDNA; 486 BP.

AC AAT32599;

DT 01-OCT-1996 (first entry)

DE P. aeruginosa OprF C-terminal region cDNA.

KW Vaccine; fusion protein; OprF; OprI; outer membrane protein;
 antibody; glutathione-S-transferase; GST; ds.

OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).

PN EP717106-A1.

PD 19-JUN-1996.

PE 17-NOV-1995; 95EP-0118098.

PR 16-DEC-1994; 94EP-0120023.

PA (BEHW) BEHRINGWERKE AG.

PI Broecker M, Domdey H, Hungerer K, Knapp B, Ulrich B;
 Von Specht BU;

DR WPI: 1996-279559/29.

DR P-PSDB; AAR99625.

PT Fusion proteins for Pseudomonas aeruginosa vaccines - contg.
 fragments of outer membrane proteins I and F.

PS Disclosure; Page 11-12; 23pp; English.

CC A cDNA sequence (AAT32599) codes for amino acids 190-350 (AAR99625) of
 the Pseudomonas aeruginosa ATCC 33354 outer membrane protein F
 (OprF), i.e. the C-terminal region of the protein. It was obt.
 by PCR amplification of the OprF gene in plasmid pFsaui (see also
 AAT32592-93). A hybrid gene between glutathione-S-transferase and
 the OprF sequence was expressed in Escherichia coli. OprF-OprI
 (AAT32600) and OprI-OprF (AAT32600) hybrid genes were constructed and
 the encoded proteins (AAR99626-27) were tested for their efficacy as
 vaccines against P. aeruginosa infection.

SQ Sequence 486 BP; 109 A; 158 C; 138 G; 81 T; 0 other;

Alignment Scores:

Pred. No.: 1,2e-11 Length: 486
 Score: 171.50 Matches: 48
 Percent Similarity: 46.15% Conservative: 18
 Best Local Similarity: 33.57% Mismatches: 67
 Query Match: 19.92% Indels: 11
 DB: 17 Gaps: 2

US-09-674-779B-2 (1-172) x AAT32599 (1-486)

QY 37 AlaProThrgIyTThrgIyValIleTThrgIyValAlaProLeuValAspAsnAsp 56
 Db 32 GTCGACGTCGACGACGACGCGGTCTGGGACAC-GTCGACAGAGTCCCGACACCCCG 90

QY 57 GluThrValIleAlaLeuAlaSerIleuProSerIleu----- 69
 Db 91 GCCACGTCACCGCTGACGCCACAGGCTGCCGCTGTGCCGAAGTCGTACCGCTACAG 150
 QY 70 -----ValTyrPheAspPheAspSerAspGluIleTyrProGluAlaIleIleu 87
 Db 151 CTGAGAGTGAAGTGCACCTTCGACCAAGTCAAGGTCMAAGACACAGCTACGTCACATC 210
 QY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValIleuValAlaGlyHis 107
 Db 211 AAGAACCTGGCCGACTCATGAAGCAGTACCGCTCCACTTCACACCGCTTGAAGTCAT 270
 QY 108 ThrAspGluArgGlySerArgIuTyrAsnMetSerIleuGlyGluArgArgAlaValAla 127
 Db 271 ACCGACTCCGTGCGTACGACGCTTACACACGAAAGCTGCCGACGTCGCCAACGCC 330
 QY 128 ValArgAsnTyrLeuLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleIleSer 146
 Db 331 GTTCGTGAGCTACTGTGTCAACAGTACGCTGTGGAAGTGTGCGGTACACGCTGTGCGT 390
 QY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
 Db 391 TACGGGAGTCCCGCGCGGTTGCCGACCAACGCCACCGCTGAAGGCCGCTATCAACGT 450
 QY 167 ArgAlaGlu 169
 Db 451 CGCGTTGAA 459

RESULT 30

AAT34419
 ID AAT34419 standard; cDNA; 681 BP.

AC AAT34419;

DT 01-OCT-1996 (first entry)

DE P. aeruginosa OprI-OprF hybrid gene.

KW Vaccine; fusion protein; OprF; OprI; outer membrane protein;
 antibody; glutathione-S-transferase; GST; ds.

OS Pseudomonas aeruginosa serotype 6 (ATCC 33354).

PN EP717106-A1.

PD 19-JUN-1996.

PE 17-NOV-1995; 95EP-0118098.

PR 16-DEC-1994; 94EP-0120023.

PA (BEHW) BEHRINGWERKE AG.

PI Broecker M, Domdey H, Hungerer K, Knapp B, Ulrich B;
 Von Specht BU;

DR WPI: 1996-279559/29.

DR P-PSDB; AAR99627.

PT Fusion proteins for Pseudomonas aeruginosa vaccines - contg.
 fragments of outer membrane proteins I and F.

PS Disclosure; Page 15-16; 23pp; English.

CC A cDNA sequence (AAT34419) codes for a fusion (AAR99627) between amino
 acids 21-83 of the outer membrane protein I (OprI) and amino acids
 190-350 of outer membrane protein F (OprF) of Pseudomonas aeruginosa
 ATCC 33354, the 2 moieties being separated by a dipeptide linker. A
 glutathione-S-transferase fusion with the hybrid gene was expressed
 in Escherichia coli and Saccharomyces cerevisiae transformants. The
 oprI-oprF hybrid protein was significantly less immunogenic than an
 oprI-oprI fusion (AAR99626).

Sequence 681 BP; 158 A; 221 C; 195 G; 107 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.91e-11	48	18	67	11	2
Percent Similarity:	46.15%					
Best Local Similarity:	33.57%					
Query Match:	19.92%					

US-09-674-779b-2 (1-172) x AAT34419 (1-681)

```

OY 37 AlaprotHrGlyTyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnasp 56
    |||||
DB 227 GCTCCGACCTCCGACACGACGAGGCGTCTGCCACAC-CTGCACAAAGTCCCGGACACCCG 285
OY 57 GluThrValIleAlaLeuAlaSerLysLeuProSerLeu----- 69
DB 286 GCCAACGTCACCGCTTGACGCCAACGGCTGCCGCTGCTCCGACAGTCCGACGCTACAG 345
OY 70 -----ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeu 87
DB 346 CTGACGCTGAAGTTCGCTTCGACAGTCCAGAGTCAAGAGACAGCTACGCTGACATC 405
OY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHis 107
DB 406 AAGAACCTGGCGGCTCATGATGAAGACATCCCGCTCCACCTCCACACCGCTGAAGTCAAT 465
OY 108 ThrAspGluArgGlySerArgGlyTyrAsnMetSerLeuGlyLysArgArgAlaValAla 127
DB 466 ACCGACCTCCGCTGCTGACGACGCTTACACCAAGCTGTCGACGCTGCTGCCAACGCC 525
OY 128 ValArgAsnTyrIleLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleLeuSer 146
DB 526 GTTCGTGACGACGCTGCTGACAGATGAGTGTGAAAGCTGCTGGAACGCTGTCGGT 585
OY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
DB 586 TACGGCGAGTCCCGCGGTGCGCACACGCCACCGCTGAAGGCCGCTATCAACCGT 645
OY 167 ArgAlaGlu 169
DB 646 CGCGTTGAA 654

RESULT 31
AAN82023
ID AAN82023 standard; DNA; 1253 BP.
XX
AC AAN82023:
XX
DT 22-OCT-1990 (first entry)
XX
DE Outer membrane protein F of Pseudomonas aeruginosa.
XX
KW Outer membrane protein F; OMP; vaccination; antibodies; ss.
XX
OS Pseudomonas aeruginosa.
XX
FT Key Location/Qualifiers
FT CDS 64..1114
FT /tag=a
FT /product=Outer membrane protein F
XX
DE3718591-A.
XX
PD 15-DEC-1987.
XX
PF 03-JUN-1987; 87DE-3718591.
XX
PR 03-JUN-1987; 87DE-3718591.
XX
PA (BEHW ) BEHRINGWERKE AG.
XX

```

PI Domdey H, Lottspeich F, von Specht B-U, Duchene M;
 XX WPI: 1988-361619/51.
 DR P-PSDB; AAP82053.

PT New outer membrane protein F of Pseudomonas aeruginosa -
 PT DNA sequences encoding it and derived antibodies, useful for
 PT vaccination and diagnosis.
 PS Disclosure; p; German.

XX The protein is isolated from the OMP of P. aeruginosa serotype 6
 CC ATTC 33354) and purified by HPLC. The amino-terminal and trypsin
 CC fragments are sequenced and a series of oligonucleotide probes
 CC constructed corresponding to the established sequences. These probes are
 CC used to screen a gene bank of 15-20 kb fragments of genomic DNA in
 CC lambda Emu 3. One positive clone includes a 15 kb insert contg. the
 CC protein gene, which can be isolated as a 2.5 kb PetI fragment. This
 CC fragment cannot be cloned into a high copy no. vector because of the
 CC toxicity of the gene prod., so is subcloned as two fragments with an
 CC overlapping region of about 500bp. Ab's are raised by usual immunisation
 CC or cell-fusion procedures. The DNA is useful in diagnosis.

Sequence 1253 BP; 279 A; 389 C; 358 G; 227 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	4.47e-11	1253	56	27	79	22
Percent Similarity:	171.50%					
Best Local Similarity:	45.36%					
Query Match:	30.60%					

US-09-674-779b-2 (1-172) x AAN82023 (1-1253)

```

OY 8 AlaAlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGly-----CysAla 24
    :|||
DB 542 ACTACGGCTTGAGAGAGGCTGACACAGTCCACAGGATGATGCTGCTGGCTGGCG 601
OY 25 AsnLysSerThrSerGlnVal-----MetValAlaProAsn 36
    :|||
DB 602 TCGGCTTCAACTTGGTGTGGAAGCGCTCCGCTCCGGAACCGGTGCGACGTT 661
OY 37 AlaprotHrGlyTyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsn 56
    |||||
DB 662 GCTCCGACCTCCGACACGACGCGCTGTCGACAC-CTGCACAACTGCCGACACCCG 720
OY 57 GluThrValIleAlaLeuAlaSerLysLeuProSerLeu----- 69
DB 721 GCCAACGTCACCGCTGACGCCAACGCTGCCGCTGTCGCCGAAGTCCGACGCTACAG 780
OY 70 -----ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeu 87
DB 781 CTGACGCTGACGCTGCTGACATTCACAAAGTCCAAAGAACGCTGACATC 840
OY 88 AspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHis 107
DB 841 AAGAACCTGCGGCTCATGATGAAGACAGTACCGCTCACTCCACACCGCTGAAGCTAT 900
OY 108 ThrAspGluArgGlySerArgGlyTyrAsnMetSerLeuGlyLysArgArgAlaValAla 127
DB 901 ACCGACCTCCGCTGCTGACGACGCTTACACCAAGAGTGTGCGACGCTGCTGCCAACCC 960
OY 128 ValArgAsnTyrIleLeuGlyLys---GlyIleAsnGlnAlaSerValGluIleLeuSer 146
DB 961 GTTCGTGACGCTGCTGACAGATGAGTGTGGAAGTGTGCGTGAACGCTGCTCGT 1020
OY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166
DB 1021 TACGGCGAGTCCCGCGGTGCGCACACGCCACCGCTGAAGGCCGCTATCAACCGT 1080
OY 167 ArgAlaGlu 169

```

DB 1081 CGCGTTGAA 1089

RESULT 32

AAA38558

ID AAA38558 standard; DNA; 1319 BP.

XX AAA38558;

AC

XX 11-SEP-2000 (first entry)

DT

XX

DE Actinobacillus pleuropneumoniae ompA2 DNA.

XX

KW OmpA2; outer membrane protein; APP; pneumonic pathogen; swine;

KW Sus scrofa; serotype; antigen; vaccine; cross-reactive immunity;

KW cross-protection; diagnosis; ds.

XX

OS Actinobacillus pleuropneumoniae strain Pz420.

XX

PH Key Location/Qualifiers

FT CDS . 197..1306

FT /tag- a

FT /product- "Actinobacillus pleuropneumoniae (APP) OmpA2"

FT sig-peptide 297..253

FT /tag- b

FT mat-peptide 254..1303

FT /tag- c

XX

PN EP1001025-A2.

PD 17-MAY-2000.

XX

PF 20-OCT-1999; 99EP-0308262.

XX

PR 22-OCT-1998; 98US-0105285.

XX

PA (PF12) PFIZER PROD INC.

XX

PI Ankenbauer RG, Baarsch MJ, Campos M, Kelch RL, Rosey EL;

PI Warren-Stewart LM, Sulter BT;

XX

DR WPI: 2000-320438/28.

DR P-PSDB: AAY97900.

XX

PT Low molecular weight Actinobacillus pleuropneumoniae proteins and DNA

PT encoding them, for use as vaccines against the bacteria in swine -

XX

PS Claim 32; Page 49-51; 81pp; English.

XX

CC This sequence represents DNA encoding a low molecular weight outer

CC membrane protein, OmpA2, from Actinobacillus pleuropneumoniae (APP)

CC strain Pz420 (ATCC 98930). The invention relates to the novel APP outer

CC membrane proteins Omp20, OmpW, Omp27, OmpA1 and OmpA2 (AAV97896-Y97900)

CC and to nucleic acids encoding them (AAA38554-A38558). APP is a Gram

CC negative coccobacillus which is one of the most important swine

CC pneumonic pathogens. 12 different serotypes of APP have been recognised

CC which vary in geographic distribution. Prior art attempts at vaccinating

CC against APP have produced mainly serotype-specific immune responses. In

CC contrast, natural immunity to any one serotype seems to confer

CC significant protection from disease caused by other serotypes, suggesting

CC that natural exposure induces cross-reactive immunity to shared antigens.

CC The novel outer membrane proteins of the invention are present in all 12

CC serotypes, and may provide a target for cross-protective immunisation.

CC The novel outer membrane proteins and nucleic acids encoding them can be

CC used as a vaccine against APP in swine. They can also be used as

CC reagents for the diagnosis of APP infections.

XX

SO Sequence 1319 BP; 428 A; 252 C; 260 G; 379 T; 0 other;

Query Match: 19.74% Indels: 12

DB: 21 Gaps: 5

US-09-674-779b-2 (1-172) x AAA38558 (1-1319)

QY 34 AAlapRoasnaAlapRothGlyTYrThrGlyValIleTYrThr-----GlyVal 49

DB 824 GCTCCTGATATCCACACTGCTGTAACAGCAGGTTATCATACGTTCCGTCACAGCGCTGTA 883

QY 50 AAlapRoeValaAspasnAspGluThrValIleValaLeuAlaSerLysLeuProSerLeu 69

DB 884 GCACCAAGTTGTTGAGCAGAGAGTTGTAACRMAAAACTTCGCA-----TTCAGCTCAGAC 937

QY 70 ValIyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGlu 89

DB 938 GTTTTATTTGATTTTCGGTAAATCAAGCTTAAACACGACGACGACACAGCTTATGAC-- 994

QY 90 GlnAlaGlnPheLeuThrThrAsn-----GlnThrAlaArgValLeuValAlaGly 106

DB 995 GCACGTAACTGAAGAAATCGCTAACTTAGTTAGCTTACCACTCCAGCTATCCAAAGTTAACGT 1054

QY 107 HisThrAspGluArgGlySerArgGluThrAsnMetSerLeuGlyGluArgAlaVal 126

DB 1055 TATACAGACCCCTATCGGTAAGAAAGACTTCAAACTTACAAACCGCGTCAGAA 1114

QY 127 AlAValArgAsnTYrLeuLeuGlyLysGlyTlleAsnGlnAlaSerValGluIleIleSer 146

DB 1115 ACTGTAGCTACTACTTATGTTTCTTAAGAGCTAAACCTGCATAACGTAAGTGCATGAGT 1174

QY 147 PheGlyGluGluArgProIleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArg 166

DB 1175 TAGGTGAGCAAAACCCAGTACCGGCGCAACATGATCAAAAGT-----AAAGTGTCT 1228

QY 167 ArgAla 168

DB 1229 AAAGCA 1234

RESULT 33

AAC79664

ID AAC79664 standard; DNA; 1110 BP.

XX AAC79664;

AC

XX 08-FEB-2001 (first entry)

DT

XX

DE Virulence gene #71.

XX

KW Virulence gene; antibacterial; vaccine; bacterial infection;

KW septicemia; bronchopneumonia; rhinitis; wound infection; ss.

XX

OS Actinobacillus pleuropneumoniae.

XX

PN WO200061724-A2.

XX

PD 19-OCT-2000.

XX

PF 06-APR-2000; 2000WO-US09218.

XX

PR 09-APR-1999; 99US-0128689.

PR 10-SEP-1999; 99US-0153453.

XX

PA (PHAA) PHARMACIA & UPJOHN INC.

XX

PI Lowery DE, Fuller TE, Kennedy MJ;

XX

DR WPI: 2000-647422/62.

DR P-PSDB: AAB44589.

XX

PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence

PT genes, useful as a live attenuated vaccine against bacterial infections

XX

PS Claim 7; Pages 306-308; 322pp; English.

XX The family Pasteurellaceae encompasses several pathogens that infect a
 CC wide variety of animals. The present invention relates to virulence genes
 CC from Pasteurellaceae. The present sequence is one such virulence gene.
 CC The present sequence may be mutated in order to produce an inactive gene.
 CC The inactive virulence gene may in turn be used to produce a vaccine,
 CC which is useful for treating bacterial infections such as septiciemias,
 CC bronchopneumonias, rhinitis and wound infections.

XX
 XX
 SQ Sequence 1110 BP; 365 A; 229 C; 220 G; 296 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	7.63e-11	1110	47	16	58	10
Percent Similarity:	169.00					
Best Local Similarity:	48.09%					
Query Match:	19.63%					
DB:	21					4

US-09-674-779b-2 (1-172) x AAC79664 (1-1110)

QY 34 AlapProasnaIapProthGlyTyrThrValIleTyrThr-----GlyVal 49
 |||||
 DB 628 GCTCCTGATATCCACTCTGTACAGCAGGTTTATCATACGTTTCGCTCAAGCGCTGTA 687
 50 AlapProleuValaspasnaSpGluThrValIleValaleuAlaSerLysLeuProSerLeu 69
 |||||
 DB 668 GCACACGTTGTGAGCCAGAGTTGTACTATAAAACCTTGC-----TTCAGCTCAGAC 741
 70 ValIyrPheaspPheaspSeraspGluIleLysProGlnAlaAlaIleLeuaspGlu 89
 |||||
 DB 742 GTTATTATGATTGTTGCGTAATCAAGCTTAACACACAGCAGCAGCTTATGAC--- 798
 90 GlnIagInPheLeuThrThrAsn-----GlnThrAlaArgValIleValAlaGly 106
 |||||
 DB 799 GCAGCTAACACTGAATCGCTAACTTACGTTTACCACTCCAGCTATTCAGTTAACGGT 858
 107 HisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaVal 126
 |||||
 DB 859 TATACAGACCGTATCGGTAAAGAGCTTAAACCTTAAACCTTTCACAACGCCGTCAGAA 918
 127 AlaValaArgAsnTyrIleLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSer 146
 |||||
 DB 919 ACTGTAGCTACTACTTATGTTCTTAAAGCTCAAAACCTGCAGAACGTAACTGCAGTAGGT 978
 147 PheGlyGluGluArgProIleAlaPheGlyThr 157
 |||||
 DB 979 TACGCTGAAGCAACCCAGTAAACCGCGGCACACA 1011

RESULT 34
 AAQ78916
 ID AAQ78916 standard; DNA; 1720 BP.
 AC AAQ78916;
 XX 09-AUG-1995 (first entry)
 DE Non-typable Haemophilus influenza (NTHI) fimbria gene.
 XX Fimbria protein; vaccine; otitis media; ss.
 KW Haemophilus influenza strain 1128.
 OS
 XX Key Location/Qualifiers
 FH 406..1485
 FT CDS
 FT RBS 385..388
 FT stem_loop 1518..1547
 FT /*tag= c
 XX W09426304-A.

PD 24-NOV-1994.
 XX
 PF 12-MAY-1994; 94MO-US05477.
 XX
 PR 18-MAY-1993; 93US-0065442.
 XX
 PA (OHIO-) OHIO STATE RES FOUND.
 XX
 PI Bakaletz LO, Kolattukudy PE, Strakova T;
 DR WPI: 1995-006359/01.
 DR P-PSDB; AAR66294.
 XX
 PT Vaccine comprising non-typable Haemophilus influenza fimbria
 protein - useful in studying, preventing or reducing the
 severity of otitis media, also fimbria protein and DNA.
 XX
 PS Claim 12; Fig 5; 45pp; English.

XX
 XX The fimbria proteins from 15 randomly selected type b and non-
 CC typable clinical isolates of Haemophilus influenza share common
 CC epitopes. Thus fimbria isolated from non-typable Haemophilus
 CC influenzae 1128 strain is a particularly suitable immunogen to
 CC protect against the different non-typable H₂ influenzae that cause
 CC otitis media. Fimbria protein is produced by culturing a transformed
 CC microbial host, pref. E.coli, sporodoptera frugiperda or a mucosal
 CC pathogen. Fimbria protein (FP) produced by this process is claimed.
 CC The FP protein migrates in polyacrylamide gels to a posn. equiv. to
 CC a mol. wt. of 25.5 kd or 37.5 kd.

XX
 SQ Sequence 1720 BP; 571 A; 301 C; 343 G; 505 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	3.76e-10	1720	47	26	63	21
Percent Similarity:	165.50					
Best Local Similarity:	46.50%					
Query Match:	29.94%					
DB:	16					4

US-09-674-779b-2 (1-172) x AAQ78916 (1-1720)

QY 35 ProasnaIapProthGlyTyr-----ThrGlyValIleTyr 46
 |||||
 DB 1000 CCMAATACCGCAATTAATCAACCTTGATGTTGATCAATCGGATATTTCTTAC 1059
 47 -----ThrGlyValaIapProleuValaspasnaSpGluThrValIleValaleuAla 63
 |||||
 DB 1060 CGTTGGTGCATAGGCGAGCAGCAGCTGTGTGAGACAGCTGAATGTAAAGCAAACTTTC 1119
 64 SerLysLeuProSerLeuValTyrPheaspPheaspSeraspGluIleLysProGlnAla 83
 |||||
 DB 1120 AGC---TTAAATCTGATGTAATCTTCCGATTTGTTAAAGCAAACTTAAACCTCAACA 1176
 84 AlaAlaIleLeuaspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValIle 103
 |||||
 DB 1177 CAAGCTACATTAAGACGCGCTATGCGCAAAATTTCAACAGTTAAAGTCGAAAGTACCT 1236
 104 ValaIagIHisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArg 123
 |||||
 DB 1237 GTTGGCTGTTACACTAACCGTATGTTGTGTGACGGGTTCAACAGTAACTTCTCAACA 1296
 124 ArgAlaValaIaValaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValaIle 143
 |||||
 DB 1297 CGTGCAGATTCAGTACTACTTGTGTTGTTAAAGGTGTGACGACGACGCAATCTCA 1356
 144 IleIleSerPheGlyGluGluArgProIleAlaPheGlyThrAsnGluGlu----- 160
 |||||
 DB 1357 GCAAAGCTTACGTTAGCAACCAACCGTAACTGCGGCAACTGTGACCAAGTTAAAGCT 1416
 161 -----AlatPserGlnAsnArgArgAlaGluIleSer 171
 |||||
 DB 1417 CGTAAAGCACTTATGCTTGTCTGCTCCAGACCGTGTGTAGAAATCGCA 1467

RESULT 35
 AAL46503
 ID AAL46503 standard; DNA; 898 BP.
 XX
 AC AAL46503;
 XX
 DT 19-JUL-2002 (first entry)
 XX
 DE M catarrhalis MCA100854 gene SEQ ID NO: 21.
 XX
 KM Moraxella; vaccine; respiratory tract infection; antiinflammatory;
 KM auditory; antibacterial; otitis media; sinusitis; pneumonia; gene; ds.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200218595-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 28-AUG-2001; 2001WO-CA01221.
 XX
 PR 28-AUG-2000; 2000US-228294P.
 PR 28-AUG-2000; 2000US-228295P.
 PR 28-AUG-2000; 2000US-228296P.
 PR 29-AUG-2000; 2000US-228438P.
 PR 29-AUG-2000; 2000US-228439P.
 PR 29-AUG-2000; 2000US-228440P.
 PR 29-AUG-2000; 2000US-228441P.
 PR 29-AUG-2000; 2000US-228442P.
 PR 29-AUG-2000; 2000US-228443P.
 PR 29-AUG-2000; 2000US-228511P.
 PR 29-AUG-2000; 2000US-228512P.
 PR 29-AUG-2000; 2000US-228742P.
 PR 01-SEP-2000; 2000US-228773P.
 PR 01-SEP-2000; 2000US-229465P.
 PR 01-SEP-2000; 2000US-229474P.
 PR 01-SEP-2000; 2000US-229475P.
 PR 01-SEP-2000; 2000US-229478P.
 PR 05-SEP-2000; 2000US-229740P.
 PR 05-SEP-2000; 2000US-229803P.
 PR 05-SEP-2000; 2000US-229804P.
 PR 05-SEP-2000; 2000US-229805P.
 PR 05-SEP-2000; 2000US-229806P.
 PR 05-SEP-2000; 2000US-229809P.
 PR 05-SEP-2000; 2000US-229811P.
 PR 06-SEP-2000; 2000US-230214P.
 PR 06-SEP-2000; 2000US-230250P.
 PR 06-SEP-2000; 2000US-230252P.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 XX
 PI Loomore S, Wang J, Bradley B, Ochs M, Yang Y;
 XX
 DR WPI: 2002-401721/43.
 DR P-PSDB; AAO17571.
 XX
 PT Moraxella polypeptide and polynucleotides useful as vaccine for
 PT immunizing a host e.g. humans against disease e.g. otitis media,
 PT pneumonia, caused by infection of the bacteria
 XX
 PS Claim 2; Fig 21; 277pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC proteins from Moraxella catarrhalis. These can be used to produce
 CC vaccines which protect against M. catarrhalis infection, which can cause
 CC otitis media, respiratory infection, sinusitis, and pneumonia. The
 CC present sequence is a coding sequence of the invention.
 XX
 SQ Sequence 898 BP; 290 A; 171 C; 182 G; 255 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.75e-10 Length: 898

Score: 165.00
 Percent Similarity: 46.72%
 Best Local Similarity: 30.66%
 Query Match: 19.16%
 DB: 24
 Gaps: 1
 Matches: 42
 Conservative: 22
 Mismatches: 71
 Indels: 2
 US-09-674-779b-2 (1-172) x AAL46503 (1-898)
 QY 33 ValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGlyValAlaProIleu 52
 DB 302 GTTGCCGATAGTATTGAAACACGCCGCTGCTCTATGTTGCTTGGCT-----GATACG 355
 QY 53 ValAspAsnAspGluThrValValAlaLeuAlaSerIysLeuProSerIleValTyrPhe 72
 DB 356 GTTGAGAAATGAGATGATATTTTAAATCATGATTAATTAATTAATTAATTAATTAATTA 415
 QY 73 AspPheAspSerAspGluIleIysProGlnAlaAlaIleLeuAspGluGlnAlaGln 92
 DB 416 GCTTTAGACTCAACCGAAATTCGCCAGAAATTAAGAAATTCGATTTGGCTGCCGAA 475
 QY 93 PheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGly 112
 DB 476 AAATTAAGCGACGCTGCGTGAACCACTTGGCTATCATGTCATACGACACTCAAGCG 535
 QY 113 SerArgGluTyrAsnMetSerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeu 132
 DB 536 ACGCATGAGTATTAATCAAGATTTCAGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595
 QY 133 LeuGlyIysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgPro 152
 DB 596 GTATCAAAAGSGTGTGCTGCTGACGTTTGAACCTCAAGTGAAGTGAAGTGAAGTGAAGTGA 655
 QY 153 IleAlaPheGlyThrAsnGluAlaIleTyrSerGlnAsnArgAlaGlu 169
 DB 656 GTTGCATCAAAATGCTACCGAACAAGTGTCTCCAAACCGTGTATTGAG 706
 RESULT 36
 AAF28548
 ID AAF28548 standard; DNA; 96109 BP.
 XX
 AC AAF28548;
 XX
 DT 04-APR-2001 (first entry)
 XX
 DE Genomic fragment #35.
 XX
 KM Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KM bronchopulmonary; endocarditis; meningitis; ss.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200078968-A2.
 XX
 PD 28-DEC-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16649.
 XX
 PR 18-JUN-1999; 99US-0140121.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lagace RE, Patterson C, Berg KL;
 XX
 DR WPI: 2001-041427/05.
 XX
 PT Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -
 XX
 PS Claim 1; Page 345-368; 545pp; English.
 XX
 CC The present invention relates to a Moraxella catarrhalis genomic library

comprising of a combination of 41 nucleic acid molecules (see AAF28514-AAF28554). The library has a number of genes described in the specification e.g. is useful for identifying diagnostic and therapeutic compositions. *M. catarrhalis* (Branhamella catarrhalis) is a large aerobic, gram-negative diplococcus, normally found among the bacterial flora of human upper airways. *M. catarrhalis* is known to cause acute, localised infections such as otitis media, sinusitis and bronchopulmonary infection and life-threatening, systemic diseases including endocarditis and meningitis.

Sequence 96109 BP; 28783 A; 18910 C; 20341 G; 28075 T; 0 other;

Alignment Scores:

Pred. No.: 1.17e-07 Length: 96109
Score: 165.00 Matches: 42
Percent Similarity: 46.72% Conservative: 22
Best Local Similarity: 30.66% Mismatches: 71
Query Match: 19.16% Indels: 2
DB: Gaps: 1

US-09-674-779B-2 (1-172) x AAF28548 (1-96109)

```
OY 33 ValAlaProAsnAlaProThrGlyTyrThGlyValIleTyrThrGlyValAlaProLeu 52
    |||||  ::::|  |||  ::::|  |||
DB 68342 GTTCCGATGATGTAAGAACGCCGCTGCTATGCTGCTTTGGT-----GATACG 68395
OY 53 ValAspAsnAspGluThrValIleValAlaLeuAlaSerIleuProSerLeuValTyrPhe 72
    |||||  ::::|  ::::|  ::::|  ::::|
DB 68386 GTTGAAGAAATGAGATGATTTTATCATCATTAATACCCAAATCATTAACCTT 68455
OY 73 AspPheSerAspGluIleValysProGlnAlaAlaIleLeuAspGluGlnAlaGln 92
    |||||  |||||  ::::|  |||||  |||||
DB 68456 GCTTTAGACCTACACCAAAATTCCTGATTTGGCTGCCGAA 68515
OY 93 PheLeuThrThrAsnGlnThrAlaArgValLeuValaGlyHisThrAspGluArgGly 112
    |||  |||  ::::|  ::::|  |||||  |||||
DB 68516 AAATTAAGGACAGTCCGCTGAACAACCTTGCATTCATGCTCTACGACACTCAAGGC 68575
OY 113 SerArgGluTyrAsnMetSerLeuGlyGluArgGluValAlaValaArgAsnTyrLeu 132
    ::::|  |||||  |||  |||  |||||  |||||
DB 68576 ACGCATGAGTATATCAAGATTTATCAAGATCTGCTGCTGCTGTTAAAGATATTG 68635
OY 133 LeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgPro 152
    ::::|  |||||  ::::|  ::::|  ::::|  ::::|
DB 68636 GTATCAAAAGGTGTGCTGCTGAACGTTTGAAACATCAAGGTGCAAGTTTGATATCA 68695
OY 153 IleAlaPheGlyThrAsnGluGluAlaIlePheSerGlnAsnArgArgAlaGlu 169
    ::::|  |||||  |||||  |||||  |||||  |||||
DB 68696 GTTGCATCAAAATGCTACCGAACAAGGTGCTTCCAAACCGCTGATTGAG 68746
```

RESULT 37

ABA91419 ID ABA91419 standard; DNA; 1035 BP.

AC ABA91419;

DT 18-APR-2002 (first entry)

DE Haemophilus paragallinarum strain 2403 haemagglutinin *haga* gene.

KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;

KW *haga*; gene; ds.

OS Haemophilus paragallinarum.

XX Key location/Qualifiers

FT CDS 1..1035

FT /tag= a

FT /product= "haemagglutinin"

FT sig_peptide 1..63

FT /tag= b

FT mat_peptide 64..1032

FT /tag= c

XX WO200204485-A1.

PN 17-JAN-2002.

PD 06-JUL-2001; 2001WO-AU00822.

PF 07-JUL-2000; 2000AU-0008652.

PR (UYOU) UNIV QUEENSLAND.

XX Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;

XX WPI: 2002-154917/20.

DR P-PsDB; AAM50718.

XX New haemagglutinin polypeptide isolated from *Haemophilus paragallinarum*

PT useful as a vaccine for immunising chickens against coryza caused by

PS the species

XX Claim 11; Fig 5; 67pp; English.

The present sequence is that of the coding region of the *haga* gene of *Haemophilus paragallinarum* strain 2403 (serovar A). The gene was isolated from chromosomal DNA by PCR amplification. It encodes haemagglutinin (see AAM50718). The invention provides recombinant haemagglutinin polypeptides (see AAM50716-27) and encoding nucleic acids (see ABA91417-27) from 11 strains (serovars A, B and C) of *H. paragallinarum*, the causative agent of infectious coryza of chickens. The polypeptides are useful in vaccines for immunisation against infectious coryza, as are the nucleic acids when expressed in attenuated bacteria, especially *Salmonella* or *Mycoplasma* (claimed). The recombinant polypeptide is preferably the mature protein, or a biologically active fragment, variant or derivative, that is capable of eliciting an immune response, providing protection against one or more strains of *H. paragallinarum* in chickens. Also claimed are methods of using the haemagglutinin polypeptides and nucleic acids for detection and diagnosis of infectious coryza in chickens.

XX Sequence 1035 BP; 319 A; 186 C; 234 G; 296 T; 0 other;

Alignment Scores:

Pred. No.: 5.71e-10 Length: 1035
Score: 161.50 Matches: 51
Percent Similarity: 46.06% Conservative: 25
Best Local Similarity: 30.91% Mismatches: 68
Query Match: 18.76% Indels: 21
DB: Gaps: 5

US-09-674-779B-2 (1-172) x ABA91419 (1-1035)

```
OY 22 GlyCysAlaAsnLysSerThrSerGluValMetValAlaProAsnAlaProThrGlyTyr 41
    |||  |||  |||||  |||||  |||||  |||||
DB 541 GGCAGTGTGGAAGAAATGATGAGCGGTGATATATACCAAGATCGTCTCTAATC 600
OY 42 ThrGlyValIleTyrThr-----GlyValAlaProLeuValAlaAspAspGluThrVal 59
    |||||  |||  |||||  |||||  |||||  |||||
DB 601 GCTGCTTATCTTACCGCTTTGGTCAAGTGCACAGTGTGACCTTAAGTTGTGCA 660
OY 60 LysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIle 79
    |||  |||  |||  |||  |||  |||  |||  |||
DB 661 AAAACATTTGCA-----TTAAATTCAGATGTTACTTGGCATTTGGTAAACCAATTGA 714
OY 80 LysProGlnAlaAlaIleLeuAsp-----GluGlnAlaGlnPheLeuThr 95
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
DB 715 CGTCCGAAGCAACAATGATATAGACGGTATTTATGTCGATGACACAGTTA----- 768
OY 96 ThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlu 115
    ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
DB 769 -----AAATCACTACCAAGATGATGTTGCTGTTATATACCGATTTGGTAGCAAGCA 822
OY 116 TyrAsnMetSerLeuGlyGluArgGluValAlaValAlaArgAsnTyrLeuLeuGlyLys 135
```



```

FT      mat_peptide      /tag= b
FT      64..1032
FT      /tag= c
XX      WO200204485-A1.
XX      17-JAN-2002.
XX      06-JUL-2001; 2001WO-AU00822.
XX      07-JUL-2000; 2000AU-0008652.
XX      (UYOU ) UNIV QUEENSLAND.
XX      Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX      MPI; 2002-154917/20.
XX      P-PSDB; AAM50722.
XX      New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX      PT useful as a vaccine for immunising chickens against coryza caused by
XX      PT the species
XX      PS Claim 11; Fig 5; 67pp; English.
XX      The present sequence is that of the coding region of the haga gene
XX      CC of Haemophilus paragallinarum strain 2671 (serovar B). The gene
XX      CC was isolated from chromosomal DNA by PCR amplification. It
XX      CC encodes haemagglutinin (see AAM50722). The invention provides
XX      CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX      CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX      CC A, B and C) of H. paragallinarum, the causative agent of infectious
XX      CC coryza of chickens. The polypeptides are useful in vaccines for
XX      CC immunisation against infectious coryza, as are the nucleic acids
XX      CC when expressed in attenuated bacteria, especially Salmonella or
XX      CC Mycobacteria (claimed). The recombinant polypeptide is preferably
XX      CC the mature protein, or a biologically active fragment, variant or
XX      CC derivative, that is capable of eliciting an immune response,
XX      CC providing protection against one or more strains of H.
XX      CC paragallinarum in chickens. Also claimed are methods of using the
XX      CC haemagglutinin polypeptides and nucleic acids for detection and
XX      CC diagnosis of infectious coryza in chickens.
XX      SQ Sequence 1035 BP; 318 A; 187 C; 234 G; 296 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 5, 71e-10 Length: 1035
XX      Score: 161.50 Matches: 51
XX      Percent Similarity: 46.06% Conservative: 25
XX      Best Local Similarity: 30.91% Mismatches: 68
XX      Query Match: 18.76% Indels: 21
XX      DB: 24 Gaps: 5
XX
XX      US-09-674-779B-2 (1-172) x ABA91423 (1-1035)
XX      QY 22 GlycylalanylserineSerineValineValineProlineAlaProthionineGlyTyr 41
XX      DB 541 GGACGCTGGAAAGATGGTACCGCTGATTTATACAAACACGAGTGTCTGTAAC 600
XX      QY 42 ThreonylValineIleTyrosine-----GlyValineAlaProlineValineAsparagineGluThrVal 59
XX      DB 601 GCTGGTTATCTTACCGCTTTGGTCAAGTGCACACAGTGTGTGAACCTAAGTGTGCA 660
XX      QY 60 LysalalalalaserineLeuProlineValineTyrosineAsparagineSerineAsparagineGluThr 79
XX      DB 661 AAAACATTGGCA-----TTAAATTCAGATGTTCTTGGCATTTGGTAAAGCAAAATTTA 714
XX      QY 80 LysproGlnAlaAlaAlaIleLeuAsp-----GluGlnAlaGlnPheLeuThr 95
XX      DB 715 CGTCCAGAACGACAAATATGTTAGACGTTATTTATGTTGAATCGACAGTTA----- 768
XX      QY 96 ThrasangInThrAlaArgValineValineAlaGlyHisThrAspGluArgGlySerArgGlu 115

```

```

DB      769 -----AAATCACTACAGTAGATGTTGCTGTTATCTACCGCATTTGTAGCGAAGCA 822
QY      116 TyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyrLeuGlyLys 135
DB      823 GCCAACTGAAATATATACAGACGCTGCTGATACAGTGTGCTAACTACTAGTTCTAAA 882
QY      136 GlyTLeuAsnGlnAlaSerValGluIleIleSerPheGlyGluArgProlineAlaPhe 155
DB      883 GGTGTGCTCAAGAGATGATTTCTCAACAGGTTATGTTGAACGACCAAGCAAGTGTGT 942
QY      156 GlyThrasnglu-----GluAlaTrpSerGlnAsnArg 166
DB      943 GCGAATGTGATACGGTTAAAGTCCGAAGCATTTATGCTTTTACAGACGATCGT 1002
QY      167 ArgAlaGluLeuSer 171
DB      1003 CGTGTACAAATCTCA 1017
XX      RESULT 40
XX      ABA91425
XX      ID ABA91425 standard; DNA; 1035 BP.
XX      AC ABA91425;
XX      DT 18-APR-2002 (first entry)
XX      DE Haemophilus paragallinarum strain H-18 haemagglutinin haga gene.
XX      KW Haemagglutinin; antigen; vaccine; immunisation; coryza; chicken;
XX      KW haga; gene; ds.
XX      OS Haemophilus paragallinarum.
XX      FH Key Location/Qualifiers
XX      FT 1..1035
XX      FT CDS
XX      FT sig_peptide 1..63
XX      FT mat_peptide /tag= b
XX      FT 64..1032
XX      FT /product= "haemagglutinin"
XX      FT /tag= c
XX      PN WO200204485-A1.
XX      PD 17-JAN-2002.
XX      PF 06-JUL-2001; 2001WO-AU00822.
XX      PR 07-JUL-2000; 2000AU-0008652.
XX      PA (UYOU ) UNIV QUEENSLAND.
XX      PI Terry TD, Tseng H, Hobb RI, Jennings MP, Downes J;
XX      MPI; 2002-154917/20.
XX      DR P-PSDB; AAM50724.
XX      PT New haemagglutinin polypeptide isolated from Haemophilus paragallinarum
XX      PT useful as a vaccine for immunising chickens against coryza caused by
XX      PT the species
XX      PS Claim 11; Fig 5; 67pp; English.
XX      The present sequence is that of the coding region of the haga gene
XX      CC of Haemophilus paragallinarum strain H-18 (serovar C). The gene
XX      CC was isolated from chromosomal DNA by PCR amplification. It
XX      CC encodes haemagglutinin (see AAM50724). The invention provides
XX      CC recombinant haemagglutinin polypeptides (see AAM50716-27) and
XX      CC encoding nucleic acids (see ABA91417-27) from 11 strains (serovars
XX      CC A, B and C) of H. paragallinarum, the causative agent of infectious
XX      CC coryza of chickens. The polypeptides are useful in vaccines for
XX      CC immunisation against infectious coryza, as are the nucleic acids
XX      CC when expressed in attenuated bacteria, especially Salmonella or

```




GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 6, 2003, 12:16:20 ; Search time 1402 Seconds

(without alignments)
1986.893 Million cell updates/sec

Title: US-09-674-779b-2

Perfect score: 861

Sequence: 1 MMHIIQIAAAALSVLFEM.....IAFGTNEAMQNRRAELSY 172

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delpop 6.0 , Delpext 7.0	

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xl
-Q/cgr2.1/USPTO.spool/US0967477/runat_06072003_121614_9813/app.query.fasta_1.327
-DB=EST -OFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -IOPTCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=100
-DOCALIGN=200 -THR.SCORE=dot -THR.MAX=100 -THR.MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US09674779.acgn_1.1.1525 -etunat_06072003_121614_9813 -NCPU=6 -ICPU=3
-NO_MAP -LARGEOUTER -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DELTIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estinu: *
4: em_estinu: *
5: em_estlov: *
6: em_estlpl: *
7: em_estro: *
8: em_nlc: *
9: gb_estl: *
10: gb_estl2: *
11: gb_hlc: *
12: gb_est4: *
13: gb_est5: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vit: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	206	23.9	747	10	BE530264	BE530264 M7K09STM
2	182.5	21.2	779	17	BH392144	BH392144 AG-ND-131
3	163	18.9	236	12	BE635668	BE635668 N106A12D
4	163	18.9	246	12	BG451732	BG451732 NF092B07D
5	163	18.9	305	12	BG451173	BG451173 NF092B07D
6	161.5	18.8	367	17	AQ990953	AQ990953 RIC01801
7	150	17.4	567	10	BE248360	BE248360 NF003G05D
c 8	130	15.1	801	17	BH404849	BH404849 AG-ND-126
9	126	14.6	603	17	BH403497	BH403497 AG-ND-160
10	124.5	14.5	814	17	BH386478	BH386478 AG-ND-170
11	124	14.4	512	17	BH374389	BH374389 AG-ND-177
12	121	14.1	658	17	BH387230	BH387230 AG-ND-169
13	120	13.9	822	17	BH393516	BH393516 AG-ND-160
14	117.5	13.6	743	17	BH389388	BH389388 AG-ND-132
15	117	13.6	448	17	BH377553	BH377553 AG-ND-155
16	113.5	13.2	800	17	BH376004	BH376004 AG-ND-135
17	105.5	12.3	704	17	AQ012123	AQ012123 8327H1B83
18	99	11.5	649	17	BH393822	BH393822 AG-ND-169
c 19	93.5	10.9	296	9	A1904113	A1904113 IL-BT042-
c 20	86.5	10.0	1152	14	BM802521	BM802521 AGENCOURT
c 21	82.5	9.6	686	13	B1290925	B1290925 B1290925
22	82	9.5	638	17	BH388977	BH388977 AG-ND-126
23	81.5	9.5	594	13	BM499884	BM499884 952035D03
24	81.5	9.5	930	14	BQ678932	BQ678932 AGENCOURT
25	81	9.4	565	13	BM448527	BM448527 DSA025B03
26	81	9.4	620	10	BE372739	BE372739 601224250
c 27	80.5	9.3	1591	12	BE729857	BE729857 601224250
c 28	80	9.3	458	12	BF758001	BF758001 QY3-CT055
29	80	9.3	908	14	BQ218323	BQ218323 AGENCOURT
c 30	80	9.3	1010	13	BM475133	BM475133 AGENCOURT
c 31	80	9.3	1131	13	BM1670158	BM1670158 603294002
c 32	79.5	9.2	385	17	AQ847628	AQ847628 LMAJFV1.1
33	79.5	9.2	480	10	BE135571	BE135571 QY3-CT055
34	79.5	9.2	919	17	CNS07510	AL430260 clone BAO
35	79	9.2	460	12	W20805	W20805 mb91c06.r1
36	79	9.2	1083	14	BG261745	BG261745 602373532
37	78.5	9.1	503	10	BE261626	BE261626 601149316
38	78	9.1	360	14	D34784	D34784 CELK012E3F
c 39	78	9.1	470	17	AQ954996	AQ954996 nbe0076M
40	78	9.1	676	13	BM491955	BM491955 p9p2n.pk0
41	77.5	9.0	412	10	BE640026	BE640026 946044H02
42	77.5	9.0	457	17	CNS07ERR	AL608834 Anopheles
43	77.5	9.0	578	13	B1595953	B1595953 949075A02
44	77.5	9.0	584	13	BM501746	BM501746 952035D03
45	77.5	9.0	662	14	B0028514	B0028514 OHH14014
46	77.5	9.0	791	10	BE514244	BE514244 601315724
47	77	8.9	413	14	D55855	D55855 H0M404C01B
48	77	8.9	551	12	BG559984	BG559984 RH122_68-
49	77	8.9	576	10	AV590145	AV590145 AV590145
50	77	8.9	595	13	B1066313	B1066313 p9fln.pk0
51	77	8.9	649	14	BQ991288	BQ991288 AGENCOURT
52	77	8.9	737	13	B1329425	B1329425 602960165
53	77	8.9	877	13	B1957118	B1957118 HVSMEH000
54	77	8.9	917	10	AW106318	AW106318 um26b06.y
c 55	77	8.9	958	13	BM475111	BM475111 AGENCOURT
56	77	8.9	1747	17	BH770900	BH770900 L1MG6463
57	76.5	8.9	482	13	B1454957	B1454957 603173530
58	76.5	8.9	559	13	BM499883	BM499883 952035D03
59	76.5	8.9	718	13	B1330404	B1330404 B1330404
c 60	76.5	8.9	773	13	B1333907	B1333907 B1333907
61	76.5	8.9	810	12	BF162117	BF162117 601768777
62	76.5	8.9	840	13	B1554962	B1554962 603236264
63	76.5	8.9	902	13	B1261918	B1261918 602953882
64	76.5	8.9	905	17	CNS06M01	AL418656 T3 end OF

Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

Plant Biology Division
The Central Research Institute

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7391
 Fax: 580 221 7380
 Email: gdmay@noble.org
 Insert Length: 246 Std Error: 0.00
 Plate: 092 row: B column: 07
 Seq primer: TCACACGAGAAACAGCTATGAC.
 Location/Qualifiers

FEATURES

source

1..246

/organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF092807DT"
 /clone_1ib="Drought"
 /tissue_type="Plantlets"
 /dev_stage="Pooled timepoints"
 /note="Vector: Lambda Zap. Contains a mixture of entire
 plantlets harvested in a series of days-post-watering
 timepoints."

BASE COUNT 67 a 56 c 61 g 61 t 1 others
 ORIGIN

Alignment Scores:

Pred. No.: 4 01e-11 Length: 246
 Score: 163.00 Matches: 31
 Percent Similarity: 75.44% Conservative: 12
 Best Local Similarity: 54.39% Mismatches: 14
 Query Match: 18.93% Indels: 0
 DB: 12 Gaps: 0

US-09-674-779b-2 (1-172) x BG451732 (1-246)

QY 116 TyraSmMetSerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLys 135

Db 3 TACAACATCTCCCTGGGTGAGACGTCTGCGAAGCCGCTTAAGATGATCCTGCAGGGTAA 62

QY 136 GlyLeaSnGlnAlaSerValGluIleIleSerPheGlyGluArgProIleAlaPhe 155

Db 63 GGGCTTCTGCAGACACGATCTCCATCTCTTACGGTAAGAAAACCTGCAGTACTG 122

QY 156 GlyThrAsnGluGluAlaTrpSerGlnAsnArgArgAlaGluLeuSerTyr 172

Db 123 GGTATGACGAGAGCGGATCTCCAAAACCGTCTGCGTACTGCTTAC 173

RESULT 5 BG451173 305 bp mRNA linear EST 16-MAR-2001

LOCUS NF099607PT1054 Drought Medicago truncatula cDNA clone NF099607DT

DEFINITION 5', mRNA sequence.

ACCESSION BG451173

VERSION BG451173.1 GI:13369962

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7381
 Fax: 580 221 7380
 Email: gdmay@noble.org
 Insert Length: 305 Std Error: 0.00
 Plate: 099 row: G column: 07
 Seq primer: TCACACGAGAAACAGCTATGAC.
 Location/Qualifiers

source

1..305

/organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF099607DT"
 /clone_1ib="Drought"
 /tissue_type="Plantlets"
 /dev_stage="Pooled timepoints"
 /note="Vector: Lambda Zap. Contains a mixture of entire
 plantlets harvested in a series of days-post-watering
 timepoints."

BASE COUNT 76 a 74 c 72 g 72 t 11 others
 ORIGIN

Alignment Scores:

Pred. No.: 5.73e-11 Length: 305
 Score: 163.00 Matches: 31
 Percent Similarity: 75.44% Conservative: 12
 Best Local Similarity: 54.39% Mismatches: 14
 Query Match: 18.93% Indels: 0
 DB: 12 Gaps: 0

US-09-674-779b-2 (1-172) x BG451173 (1-305)

QY 116 TyraSmMetSerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLys 135

Db 3 TACAACATCTCCCTGGGTGAGACGTCTGCGAAGCCGCTTAAGATGATCCTGCAGGGTAA 62

QY 136 GlyLeaSnGlnAlaSerValGluIleIleSerPheGlyGluArgProIleAlaPhe 155

Db 63 GGGCTTCTGCAGACACGATCTCCATCTCTTACGGTAAGAAAACCTGCAGTACTG 122

QY 156 GlyThrAsnGluGluAlaTrpSerGlnAsnArgArgAlaGluLeuSerTyr 172

Db 123 GGTATGACGAGAGCGGATCTCCAAAACCGTCTGCGTACTGCTTAC 173

RESULT 6 A0990953 567 bp DNA linear GSS 14-AUG-2000

LOCUS RfC01801 Photorhabdus luminescens strain W14 M13 library

DEFINITION Photorhabdus luminescens genomic clone PlG01801, DNA sequence.

ACCESSION A0990953.1 GI:9649547

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

1 (bases 1 to 567)
 French-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,
 Daborn, P.J., Bowen, D. and Blattner, F.R.
 A genomic sample sequence of the entomopathogenic bacterium
 Photorhabdus luminescens W14: potential implications for virulence
 Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
 Contact: french-constant RH
 Department of Biology and Biochemistry
 University of Bath
 South Building, Bath BA2 7AY, UK
 Tel: (44) 1225 826621
 Fax: (44) 1225 826779
 Email: bsrfc@bath.ac.uk
 This is one of 2,122 random reads from the M13 library. For
 annotation of identified clones (BLASTX, BLASTN and mapping to E.
 coli K12 genome) please see french-constant et al. 2000, Nucleic
 Acids Res.
 Seq primer: M13 Forward
 Class: shotgun.
 Location/Qualifiers

FEATURES

source

1..567

/organism="Photorhabdus luminescens"
 /strain="W14"
 /db_xref="taxon:29488"
 /clone="PlG01801"

/clone.lib="Photobabidus luminescens strain W14 M13
library"
/dev_stage="primary phase variant"
/note="Genomic DNA from strain W14 was size selected (1-2
kb) and then cloned into M13 Janus."
BASE COUNT 164 a 118 c 137 g 145 t 3 others
ORIGIN

Alignment Scores:

Pred. No.:	2,51e-10	Length:	567
Score:	161.50	Matches:	48
Percent Similarity:	50.00%	Conservative:	22
Best Local Similarity:	34.29%	Mismatches:	50
Query Match:	18.76%	Indels:	20
DB:	17	Gaps:	4

US-09-674-779b-2 (1-172) x AQ990953 (1-567)

OY 1 MetwelleuHISlegInIleAlaAlaAlaAlaAlaSerValLeuThrPheMet 20
Db 198 CTGATGTTAGCTTACCAAGTATGCGGTACAGCGTACTTCT----- 242
OY 21 ThrGlyCysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGly 40
Db 243 -----AACAGAAAT-----GCTGATTAACGATCAAGCTGCT 272
OY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeuValAspAsnGluThrValLys 60
Db 273 GTCGGCACTGTTAAGCAAACTAATCGCGGT-----CTGTACAGCAAGGAAATTCGTCGT 326
OY 61 AlaLeuAlaSerLysLeuPro-----SerLeuValTyrPheAspPheAspSerAspGlu 78
Db 327 CAGCAATATGCAAGAGCTTCAAAACACACATCGATATTGTTGTTTGCACAAATACGAC 386
OY 79 IleLysProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAspGln 98
Db 387 GTTGCGTCAAGATTTTACTCATCGCTGTATACCGCTCCTCCGCGACCAACCA 446
OY 99 ThrAlaArgValLeuValAlaGlyHisThrAspGluArgLysSerArgGluTyrAsnMet 118
Db 447 TCTTATAAGTTGTTGTGAAGAAGTCATGCTGACGACAGCTGCTCTCGATACACATC 506
OY 119 SerLeuGlyGluArgArgAlaValAlaValArgAsnTyrLeuGlyLysGlyIle 137
Db 507 GCATTANGTGAACGCCGCTGCTAATGACGTAATAATGACTGCAAGCAAAAGNGTT 564

RESULT 7

BE248360

LOCUS BE248360 237 bp mRNA linear EST 21-DEC-2000
DEFINITION NF003G05DTR1F1036 Drought Medicago truncatula cDNA clone NF003G05DT
5', mRNA sequence.

ACCESSION

BE248360
BE248360.2 GI:11964908

KEYWORDS

EST.
ORGANISM
SOURCE

barrel medic.
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 237)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
Unpublished (2000)

REFERENCE

AUTHORS

TITLE
JOURNAL

COMMENT
On Jul 13, 2000 this sequence version replaced gi:11964908.
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380

Email: gdmay@noble.org
Medicago Genome Initiative accession: MGI:S:22852
Insert Length: 866 Std Error: 0.00
Plate: 003 row: G column: 05
Seq primer: TCCACACGGAACACGCTATGAC.
FEATURES
Location/Qualifiers
1..237
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF003G05DT"
/clone_lib="Drought"
/tissue_type="Plantlets"
/dev_stage="Pooled timepoints"
/note="Vector: Lambda Zap. Contains a mixture of entire
plantlets harvested in a series of days-post-watering
timepoints."

BASE COUNT 66 a 53 c 58 g 60 t
ORIGIN

Alignment Scores:

Pred. No.:	1.78e-09	Length:	237
Score:	150.00	Matches:	31
Percent Similarity:	74.14%	Conservative:	12
Best Local Similarity:	53.45%	Mismatches:	14
Query Match:	17.42%	Indels:	1
DB:	10	Gaps:	0

US-09-674-779b-2 (1-172) x BE248360 (1-237)

OY 116 TyrAsnMetSerLeuGlyArgAlaValAlaValArgAsnTyrLeuGlyLys 135
Db 3 TACACATCTCCCTGGGTGAGCTGCGAACGCCGTTAAGATGTACTGACGAGGTAA 62
OY 136 GlyTLeuAsnGlnAlaSerValGluIleIleSerPhe-GlyGluGluArgProIleAlaP 155
Db 63 GGGGTTTTCGACAGCAGCATCTCCATGCTTTTACAGGATAAGAAAACCTGCACTACT 122
OY 155 eGlyThrAsnGluGluAlaTyrSerGlnAsnArgAlaGluLeuSerTyr 172
Db 123 GGGTCAATGACGAAGCGGCTACTTCCAAAACCGTGGCGTACTGCTTTAC 174

RESULT 8

BH404849/c

LOCUS BH404849 801 bp DNA linear GSS 11-DEC-2001
DEFINITION AG-ND-126C17.TF ND-TM Anopheles gambiae genomic clone AG-ND-126C17
, DNA sequence.

ACCESSION

BH404849
BH404849.1 GI:17351065

KEYWORDS

GSS.
SOURCE
ORGANISM

Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 801)
Shetty, J., Malek, K., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)

JOURNAL

COMMENT
Other_GSSs: AG-ND-126C17.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

Email: bjl@loftus.tigr.org
This clone is from an A. gambiae BAC library (ND-TM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center

Tel: 301 838 0208

1 (bases 1 to 814)
Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.,
Direct Submission of BAC-end sequences from *Anopheles gambiae*
Unpublished (2001)
Other GSSs: AG-ND-170A18, TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

Email: b1ofus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 For
 Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1..814
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-170A18"
 /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 275 a 141 c 184 g 214 t

ORIGIN

Alignment Scores:
 Pred. No.: 2.68e-05 Length: 814
 Score: 124.50 Matches: 29
 Percent Similarity: 52.94% Conservative: 25
 Best Local Similarity: 28.43% Mismatches: 47
 Query Match: 14.46% Indels: 1
 DB: 17 Gaps: 1

US-09-674-779b-2 (1-172) x BH386478 (1-814)

QY 70 ValTYRPhaspPheaspSerAspGluIlelyserProGlnAlaIleleuaspGlu 89

DB 410 ATCTATTTCGATTTCACAAAGCTACTATCCAGTCGATCTTATCTCAAGTAGAGCT 469

QY 90 GlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAsp 109

DB 470 GCTGCTGAGATTATCAAGAGAGCTAAAGCTAAAGCTCTTAACTCGGTACACAGAT 529

QY 110 GluArgGlySerArgGlyTyrAsnMetSerLeuGlyGluArgGalaValAlaValArg 129

DB 530 GCTAAAGGTAACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 589

QY 130 AsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSerValGluIleIleSerPheGly--- 148

DB 590 AAAGCTCTTGAAGCTAGAGGTGTAACGCTCTCAATTGAATCTGTAAAGTGTAGTCT 649

QY 149 GluGluArgProIleAlaPheGlyTyrAsnGlnGluAlaTyrSerGlnAsnArgGala 168

DB 650 AAAGAGCTACTGACACGCTACTGCTCTTACGAGGCTACACAGTAGACAGAAAGTT 709

QY 169 GluLeu 170

DB 710 GAAGTG 715

RESULT 11

BH374389 512 bp DNA linear GSS 10-DEC-2001

DEFINITION AG-ND-177A3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-177A3,

ACCESSION BH374389

VERSION BH374389.1 GI:17320531

KEYWORDS GSS

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 512)

AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submision of BAC-end sequences from Anopheles gambiae

JOURNAL COMMENT

Unpublished (2001)
 Other GSSs: AG-ND-177A3.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b1ofus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 For
 Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1..512
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-177A3"
 /note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 177 a 69 c 124 g 142 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.44e-05 Length: 512
 Score: 124.00 Matches: 29
 Percent Similarity: 54.44% Conservative: 20
 Best Local Similarity: 32.22% Mismatches: 39
 Query Match: 14.40% Indels: 2
 DB: 17 Gaps: 2

US-09-674-779b-2 (1-172) x BH374389 (1-512)

QY 54 AspAsnAspGluThrValIleAlaSerIleuProSerLeuValTyrPheasp 73

DB 234 GACAAATTAACAGGTTGCACTGACGATGAAGACGAAATTAATAAT---CTGATATTCCAT 290

QY 74 PheAspSerAspGluIlelyserProGlnAlaIleleuaspGluAlaGlnPhe 93

DB 291 TTTAATTAACCAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAAT 350

QY 94 LeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySer 113

DB 351 ATTAAGAGCAT---GGTGAATTTTCTATTATACCGACATACAGATCTTAAGGTAT 407

QY 114 ArgGluTyrAsnMetSerLeuGlyGluArgGalaValAlaValArgAsnTyrLeuLeu 133

DB 408 GCAGCATACCACTTAAGCTTTCACAGAGAAAGCGCTGCTGTGTGGAGCTCTGGA 467

QY 134 GlyLysGlyIleAsnGlnAlaSerValGlu 143

DB 468 GCTAGAGGAGAGTTCTTCCAGACTTTAA 497

RESULT 12

BH387230 658 bp DNA linear GSS 11-DEC-2001

DEFINITION AG-ND-169L12.TF ND-TAM Anopheles gambiae genomic clone AG-ND-169L12,

ACCESSION BH387230

VERSION BH387230.1 GI:17333371

KEYWORDS GSS

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 658)

AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submision of BAC-end sequences from Anopheles gambiae

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Anopheles.
1 (bases 1 to 658)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-169L12.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..658
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-169L12"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT
ORIGIN
260 a 86 c 132 g 180 t

Alignment Scores:

Pred. No.: 5.31e-05 Length: 658
Score: 121.00 Matches: 34
Percent Similarity: 44.78% Conservative: 26
Best Local Similarity: 25.37% Mismatches: 67
Query Match: 14.05% Indels: 7
DB: 17 Gaps: 2

US-09-674-779b-2 (1-172) x BH387230 (1-658)

QY 23 CysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyThr 42
DB 201 TGCCTGACGACGAAAGTGGAAATTTACTATTCTGAGAGACGACCTGAAGTTCCA 260
QY 43 GlyVal---lLeTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAla 61
DB 261 AAGGTCGAAATAATTTCTCTCCCAATGCAAAAAGATTTAAGCAGCTCGGAAAGCT 320
QY 62 LeuAlaSerLysLeuProSerLeuValTyrPheAspPheSpSerAspGluLeuLysPro 81
DB 321 GATATTGAAAT-----ATTAAATTTTGAATAAATAAAGCTCTCTTACAGCA 365
QY 82 GlnAlaAlaAlaAlaLeuAspGlnGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArg 101
DB 366 GATGGCAAAATTTGGCAGATGAGATCCTTAAGCTGTAATGAAATCTCTAAATTAAAA 425
QY 102 ValLeuValAlaGlyThrAspGluArgGlySerArgGluTyrAsn-MetSerLeuG1 121
DB 426 CTTTCTATAGAGGAGGATCTGACCAACATGCTACAAAGAACACATTAAGCAATTAAC 485
QY 121 YgluArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaLase 141
DB 486 ATTGACAGAGCTACCAACATATATATTCTTAACAGATTAAGGATTAATCTGATAG 545
QY 141 ValGluIleIleSerPheGlyGluGluArgProIleAla 154
DB 546 ACTACAAACAAAGGTTTGGACAGATTAATCTTGGCT 585

RESULT 13
BH393516 822 bp DNA linear GSS 11-DEC-2001
LOCUS BH393516
DEFINITION AG-ND-160N3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-160N3,
DNA sequence.
ACCESSION BH393516
VERSION BH393516.1 GI:17339657
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 822)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSs: AG-ND-160N3.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..822
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-160N3"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT
ORIGIN
297 a 104 c 178 g 243 t

Alignment Scores:

Pred. No.: 0.000103 Length: 822
Score: 120.00 Matches: 30
Percent Similarity: 53.33% Conservative: 18
Best Local Similarity: 33.33% Mismatches: 38
Query Match: 13.94% Indels: 4
DB: 17 Gaps: 2

US-09-674-779b-2 (1-172) x BH393516 (1-822)

QY 57 GluThrValLysAlaLeuAlaSerLysLeuProSer-----LeuValTyrPheAsp 73
DB 227 AAGCAATTAACACAGGTTGCACTGAAGTTAAACAGAAATTAATAATGCTATTTCAT 286
QY 74 PheAspSerAspGluIleLysProGlnAlaAlaAlaLeuAspGlnGlnAlaGlnPhe 93
DB 287 TTTATATAAGCAACATTAACAAATGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAT 346
QY 94 LeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyThrAspGluArgGlySer 113
DB 347 ATTAAGGCAAT---GGTGAATATTATCTTAAACCGACATACAGATCTTAAGGATAG 403
QY 114 ArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValAlaArgAsnTyrLeu 133
DB 404 GCACATTAACACCTTAAGCTTTCCAGAGAAAGACGGCTGCTGTTGTTGAGCTTGGAA 463

QY 134 GlytysgilyleasnglnalaserValglu 143
 DB 464 GCTAGAGGAGTTTCTCCAGTACTTAA 493

RESULT 14
 BH389388 743 bp DNA linear GSS 11-DEC-2001
 LOCUS AG-ND-132E3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-132E3,
 DEFINITION DNA sequence.
 ACCESSION BH389388
 VERSION BH389388.1 GI:17335529
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

REFERENCE 1 (bases 1 to 743)
 AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other_GSSs: AG-ND-132E3.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjl@fusteligr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source location/Qualifiers
 1..743
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-132E3"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1, Site_1: HindIII"

BASE COUNT 269 a 100 c 169 g 205 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.000183 Length: 743
 Score: 117.50 Matches: 26
 Percent Similarity: 55.41% Conservative: 15
 Best Local Similarity: 35.14% Mismatches: 32
 Query Match: 13.65% Indels: 1
 DB: 17 Gaps: 1

US-09-674-779b-2 (1-172) x BH389388 (1-743)

QY 70 ValtyrphaspheaspSeraspSgluUlelySPROGlnAlaAlaIleLeuaspGlu 89
 DB 275 GTGATTTCCATTATTAATGAACAATACAAATGAGTCGCAAGCTGATGTA 334

QY 90 GlnAlaGlnPheLeuThrThAsnGlnThrAlaArgValLeuValAlaGlyHisThAsp 109
 DB 335 GCTGCAACATTATTAAAGCAAT--GGTGAATATTATTAACCGGACATACAGAT 391

QY 110 GluArgGlySerArgGluTyTyrAsnMetSerLeuGlyGluArgAlaValAlaValArg 129
 DB 392 GCTAAAGTATGTCAGCATACACCTTAAGCTTTCCAGAACAGAGCGGCTGCTGTGTT 451

QY 130 AsnTyrlleuGlyleasnglnalaserValglu 143
 DB 452 GCAAGCTCTGAACTAGAGAGTTTCTCCAGTACTTAA 493

RESULT 15
 BH377553 448 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-155H3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-155H3,
 DEFINITION DNA sequence.
 ACCESSION BH377553
 VERSION BH377553.1 GI:17323695
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

REFERENCE 1 (bases 1 to 448)
 AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other_GSSs: AG-ND-155H3.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjl@fusteligr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 For
 Class: BAC ends.

FEATURES
 source location/Qualifiers
 1..448
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-155H3"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1, Site_1: HindIII"

BASE COUNT 167 a 71 c 96 g 114 t
 ORIGIN

Alignment Scores:
 Pred. No.: 9.17e-05 Length: 448
 Score: 117.00 Matches: 33
 Percent Similarity: 52.94% Conservative: 12
 Best Local Similarity: 38.82% Mismatches: 36
 Query Match: 13.59% Indels: 4
 DB: 17 Gaps: 3

US-09-674-779b-2 (1-172) x BH377553 (1-448)

QY 87 LeuaspGluGlnAlaGlnPheLeuThrThAsnGlnThrAlaArgValLeuValAlaGly 106
 DB 39 CTTGACAGATTGGCGCA---CTTATATATATATGCAATTCAGATTTT-----GGA 89

QY 107 HisThAspGluArgGlySerArgGluTyTyrAsnMetSerLeuGlyGluArgAlaVal 126
 DB 90 AATTGCGAATATTTCCGCAACTTTGCAATCAATAAAGCTGTCGAAATGCGCCAAT 149

QY 127 AlaValArgAsnTyrlleuGlyleasnglnalaserValglu---Ilelle 145
 DB 150 GCGGTTAACGATATCTCAAGTCACAAATTTGGAACAAATATATCAAACTGGAAAGTGTAGTA 209


```

Db      269 CGTACTCGACCGCCCTGCGCATCGCCGAGACGGCTGCGCGCGTGGGATGGC 328
      |||||  ||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      149 GUGUAGUProIIeAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgArgAla 168
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      329 GAACGG-----TTCTTGCGCCGACAGAAAGACCCCGCAGCGCGAAGACCGGGGTC 382
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      169 Glu 169
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      383 GAG 385

RESULT 18
BH393822
LOCUS   BH393822               649 bp    DNA        linear    GSS 11-DEC-2001
DEFINITION AG-ND-169D8.TF ND-TAM Anopheles gambiae genomic clone AG-ND-169D8,
DNA sequence.
ACCESSION BH393822
VERSION   BH393822.1  GI:17339963
KEYWORDS  GSS:
SOURCE    African malaria mosquito.
ORGANISM  Anopheles gambiae
           Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
           Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
           Anophelinae.
REFERENCE 1 (bases 1 to 649)
AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL   Unpublished (2001)
COMMENT   Other GSSs: AG-ND-169D8.TF
           Contact: Brendan J Loftus
           Department of Eukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0208
           Fax: 301 838 3543
           Email: b.loftus@tigr.org
           This clone is from an A. gambiae BAC library (ND-TAM) provided by
           F.H. Collins and sequenced by The Institute for Genomic Research
           (TIGR). The BAC library was generated from A. gambiae PEST strain
           DNA. All DNA was extracted from newly hatched first instar larvae
           to minimize the inclusion of DNA from microorganisms that inhabit
           the gut. The DNA is derived from mixed sexes of larvae. The BAC
           library was constructed at Texas A&M University BAC Center
           University, College Station, Texas 77843-2123, USA using a HindIII
           partial digest.
           Seq primer: M13 For
           Class: BAC ends.
FEATURES
     source             location/Qualifiers
     1..649
        /organism="Anopheles gambiae"
        /strain="PEST"
        /db_xref="taxon:7165"
        /clone_lib="AG-ND-169D8"
        /clone_11b="ND-TAM"
        /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT  225 a      89 c      151 g      184 t
ORIGIN
Alignment Scores:
  Pred. No.:      0.0354      Length:      649
  Score:          99.00       Matches:      24
  Percent Similarity: 54.93%   Conservative: 15
  Best Local Similarity: 33.80% Mismatches:    30
  Query Match:    11.50%      Indels:       30
  DB:             17         Gaps:         2
US-09-674-779B-2 (1-172) x BH393822 (1-649)
Qy      54 AspaenAspGluThrVallyAlaLeuAlaSerlyseuProSerLeuValTyrPheasp 73
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      234 GACATTAACCAAGTTCGACACGTGAGATACAAACAGATTAAAAAT--GTGTATTTCAT 290
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      74 PheaspSerAspGluIlelyProGluAlaAlaIleleuAspGluGluAlaGlnPhe 93

```

```

Db      291 TTTAATAAAGCAACATATCAATAATGATGAGGCAAGACTAGATGACTCCAAACAT 350
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      94 LeuThrThraSncInlhrAlaArgValLeuValAlaGlyHisThraSpLuarGlySer 113
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      351 ATTAAGAGCAAT--GGTGGAATTAATCTTAATACCGCATACAGATGCTTAAGGTAGT 407
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy      114 ArgLutyrAsnMetSerLeuGlyGluArgArg 124
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      408 GCAGCATACACACTTAAGCTTTTCACACCAAGA 440

RESULT 19
AI904113/c
LOCUS   AI904113               296 bp    mRNA        linear    EST 30-MAR-2000
DEFINITION IL-BP042-231198-008 BT042 Homo sapiens cDNA, mRNA sequence.
ACCESSION AI904113
VERSION   AI904113.1  GI:6494500
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 296)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
           Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
           Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
           Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
           M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
           Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
           sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE  20202663
COMMENT   Contact: Simpson A.J.G.
           Laboratory of Cancer Genetics
           Ludwig Institute for Cancer Research
           Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
           Brazil
           Tel: +55-11-2704922
           Fax: +55-11-2707001
           Email: asimpson@ludwig.org.br
           This sequence was derived from the FAPESP/LICR Human Cancer Genome
           Project. This entry can be seen in the following URL
           (http://www.ludwig.org.br/seq/gethtml.pl?tl=IL-BT042-008.html
           &t3=231198&t4=1)
           Seq primer: puc 18 forward.
FEATURES
     source             location/Qualifiers
     1..296
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="BT042"
        /sex="female"
        /dev_stage="Adult"
        /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
        SmaI; A mini-library was made by cloning products derived
        from ORESTES PCR (U.S. Letters Patent application No. 196
        716 - Ludwig Institute for Cancer Research) profiles
        into the puc 18 vector. Reverse transcription of tissue
        mRNA and cDNA amplification were performed under low
        stringency conditions."
BASE COUNT  43 a      88 c      97 g      68 t
ORIGIN
Alignment Scores:
  Pred. No.:      0.0491      Length:      296
  Score:          93.50       Matches:      21
  Percent Similarity: 48.19%   Conservative: 19
  Best Local Similarity: 25.30% Mismatches:    36
  Query Match:    10.86%      Indels:       8
  DB:             9         Gaps:         1
US-09-674-779B-2 (1-172) x AI904113 (1-296)

```

QY 94 LeuThrThrAsnGlnThrAlaArgValLeuAlaGlyHisThrAspGluArgGlySer 113
 ||| ||| : : : : : ||| ||| : : :
 Db 284 CTTACACCCCAACGCGGCGACAGATGTCATACCGC-CATGCCACGACGATGGCGAC 226
 QY 114 ArgGluThrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyrLeuLeu 133
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 225 CCGCAGAGCAACCGCGGTTTCGCGGCAACGTGCGAGACCATCAAGACCTACGTGTG 166
 QY 134 GlyLysGlyLeuAsnGlnAlaSerValGluIleIleSerPheGlyGluGluArgProIle 153
 ||| ||| : : : : : ||| : : : : : ||| : : : : :
 Db 165 GGCAGAGGCGCTGCGCGGCGGCAACTGTGTGATCCGTGGCGAGAGTACCAACAAACCCCTG 106
 QY 154 AlaPheGlyThrAsnGlu-----GluAlaTyrSerGlnAsnArg 166
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 105 GTGGCGTCAACATGCAAGCGCGCGGCGACGCACTGATTCAGTGCCTGGAACCTACAGG 46
 QY 167 ArgAlaGlu 169
 ||| ||| : : : : : ||| : : : : : ||| : : : : :
 Db 45 CGCCTCGAG 37

RESULT 20
 BM802521/c 1152 bp mRNA linear EST 05-MAR-2002
 LOCUS BM802521
 DEFINITION AGENCOURT_6460224 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5560157
 5', mRNA sequence.
 ACCESSION BM802521
 VERSION BM802521.1 GI:19119344
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1152)
 REFERENCE NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph. D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM1285 row: 0 column: 06
 High quality sequence start: 379
 High quality sequence stop: 585.
 Location/Qualifiers
 1. 1152
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5560157"
 /clone_1lb="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 site:1: NotI; site:2: SalI; Clone: unidirectional;
 oligo-dT primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 344 a 277 c 241 g 273 t 17 others
 ORIGIN

Alignment Scores:

Pred. No.: 3.75 Length: 1152
 Score: 86.50 Matches: 30
 Percent Similarity: 48.33% Conservative: 28
 Best Local Similarity: 25.00% Mismatches: 43
 Query Match: 10.05% Indels: 19
 Db: 14 Gaps: 5

US-09-674-779b-2 (1-172) x BM802521 (1-1152)

QY 52 LeuValAspAsnAspGluThrValIleValAlaLeuAlaSerIleuProSerLeuValTyr 71
 ||| ||| : : : : : ||| ||| : : : : : ||| : : : : :
 Db 785 TTGGTCGATCAATCAAGCTTTATTAAGTCCCTGGCGGCAATATATCAAGATGCTTC 726
 QY 72 PheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeuAspGluAla 91
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 725 ATTCATCAGAGAAAGAT-----CCATCAGCATGTGCGGCGCAAGTTCTTCA 678
 QY 92 GlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArg 111
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 677 ACAATGGCGACCTCTTCGGGAAATCTCGCCT-----CCTCGTCCCTTCGCGCAG 627
 QY 112 GlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyr 131
 ||| ||| : : : : : ||| : : : : : ||| : : : : :
 Db 626 CCACGCAATGATATTCCTTGGCAGCTG-----GCCCTCCAAATAGACCTTAC 579
 QY 132 LeuLeuGlyLysGlyLeuAsnGlnAlaSerValGlu-----IleIleSerPheGlyGlu 150
 : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 578 ATCACTGGTAAAGTCCCTTCAGCATATCTCAATTCATGCTTACGTTGGCAAGTT 519
 QY 151 ArgProIleAlaPheGlyThrAsnGlu-----GluAlaTyrSer 163
 ||| : : : : : ||| : : : : : ||| : : : : :
 Db 518 ATTCCTGTCCTTGTGTTATTCATCACTCTGCAACAAATCTTGGCCAGCGCTGAGT 459

RESULT 21
 BJ290925/c 686 bp mRNA linear EST 09-APR-2002
 LOCUS BJ290925
 DEFINITION BJ290925 Y. Ogihara unpublished cDNA library, Wh_SL Triticum
 aestivum cDNA clone whs120n03 5', mRNA sequence.
 ACCESSION BJ290925
 VERSION BJ290925.1 GI:20108010
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 1 (bases 1 to 686)
 REFERENCE Ogihara, Y. and Mural, K.
 TITLE Expressed genes in Triticum aestivum
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-I
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1. 686
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="whs120n03"
 /clone_1lb="Y. Ogihara unpublished cDNA library, Wh_SL"
 /tissue_type="seed DPA30"
 /dev_stage="Feekes' scale 11.3"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site:1: EcoRI; Site:2: XhoI; Plants were grown under
 hydroponic conditions at UC Davis, salt stressed for 12
 hours, and for 7 days, then dissected and frozen (Khanov
 in J. Dvorak Lab). Total RNA was prepared from sheath
 tissue, equal quantities of RNA were pooled from the two
 samples, polyA was purified from the pooled RNA, a cDNA
 library was made, and the cDNA clones were in vivo
 excised to give Bluescript phagemids in the T7 Close lab
 at the University of California, Riverside (Akhunov, Chin
 , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang).
 Plasmid DNA preparations and DNA sequencing were
 performed in the OD Anderson lab (all other authors)."
 BASE COUNT 138 a 233 c 181 g 132 t 2 others
 ORIGIN

partial digest.

JOURNAL Unpublished (1999)

Unpublished (1999)

COMMENT

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 952035 row: D column: 03.

FEATURES

Source

1. 594
/organism="Zea mays"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/clone_lib="952 - BMS tissue from Walbot Lab (reduced RNA)"
/tissue_type="suspension culture"
/dev_stage="mixed logarithmic and stationary growth phases"
/lab_host="DH10B"
/note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."

BASE COUNT 177 a 81 c 202 g 134 t
ORIGIN

Alignment Scores:

Pred. No.: 5.49 Length: 594
Score: 81.50 Matches: 35
Percent Similarity: 40.94% Conservative: 17
Best Local Similarity: 27.56% Mismatches: 33
Query Match: 9.47% Indels: 42
DB: 13 Gaps: 7

US-09-674-779b-2 (1-172) x BM499884 (1-594)

QY 46 TyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLys 65
DB 288 TGAATGATGCTCTCAGTATGTTGAAGAGAGAGAG----- 323
QY 66 LeuProSerLeuValTyrPheAspPheAspSerAspGluLeuLysProGlnAlaAlaAla 85
DB 324 -----TTCGATTTTGTGAGATGAGCCAAAGCAGAGAGAGCGA 365
QY 86 IleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgValLeuValAla 105
DB 366 TTGCTGATGACAACTTGTGATTCCTTCTTGAGAGAA-----TAT 407
QY 106 GlyHisThrAspGlu-----ArgLysSerArgGluTyr-----Asn 117
DB 408 GCAGACAGATGATGATGACATGCGTGTAAAGGTGG---GAATTCGAACAGCCAAAT 464
QY 118 MetSerLeuGlyGluArgAlaVal-AlaValArgAsnTyrLeuLeuGlyLysGly 137
DB 465 GAGCCTATGATGACGAATTCCTCAAAATCGAATGTATGTGTGATGAGAGATAT 524
QY 137 eAsnGlnAlaSerValGluLeuLeuSerPheGlyGluGluArgProIleAlaPheGlyThr 157
DB 525 A-----GAACACAGAGAGATTTTGTTCGCCAA 551
QY 157 Asn---GluGluAlaTyr 162
DB 552 AAACCTGATTCACAGCATGG 570
RESULT 24
BQ678932
LOCUS BQ678932 930 bp mRNA linear EST 15-JUL-2002

DEFINITION

AGENCOURT 8192156 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6260167 5', mRNA sequence.

ACCESSION

BQ678932 GI:21791611

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

1 (bases 1 to 930)
NIH-MGC <http://mgc.ncl.nih.gov/>.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.
Email: cgs@bbs.fremail.nih.gov

COMMENT

Tissue Procurement: DCTD/DRP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LMNL at:

<http://image.llnl.gov>

Plate: LRCM2418 row: n column: 08

High quality sequence stop: 714.

FEATURES

source

1. 930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="6260167"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming, directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCAGAC(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 266 a 283 c 186 g 193 t 2 others
ORIGIN

Alignment Scores:

Pred. No.: 11.6 Length: 930
Score: 81.50 Matches: 38
Percent Similarity: 42.03% Conservative: 20
Best Local Similarity: 27.54% Mismatches: 36
Query Match: 9.47% Indels: 44
DB: 14 Gaps: 8

US-09-674-779b-2 (1-172) x BQ678932 (1-930)

QY 27 SerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyr 46
DB 470 AGCAGACGACAGCAGCTGTCCAGACGCTGTGG----- 511
QY 47 ThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLysLeu 66
DB 512 ACAGATTTCAAACCTTA-----AATGATACCCCTGCACACAGACAGAGAACCC 559
QY 67 ProSerLeuValTyrPheAspPhe-----Asp 75
DB 560 CCAAGACCTACATTCCTCTTATACACAGCTACAGCTCAACAACTAGTACAAACAAT 619
QY 76 SerAspGluLeuLysProGlnAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThr 95
DB 620 AGTACTGACAGCTAACCA---GGGCTTCAATTAACAGTACAGCTGCTCACTTACACACA 676
QY 96 ThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlu 115
DB 677 ACTAGTGCACCACTAAGTTGATC-----CATCCAGATGAG----- 712

OY 116 TyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsn-TyrLeuLeuGlyLy 135
 Db 713 ---GATATATCCCTGGAGAGAGAAGGCA-----CAGTTACCTA----- 749
 OY 135 sGlyLeuAsnGlnAlaSerValGluIleIleSerPheGlyGluArgPro 152
 Db 750 -----AGTATCAACGTAACTCTCTCGCCAGCAGCAGCC 785

RESULT 25
 BM448527 565 bp mRNA linear EST 01-APR-2002
 LOCUS DS4025B03.59675 An expressed sequence tag database for the
 DEFINITION halotolerant green alga, Dunaliella salina Dunaliella salina cDNA
 clone DS4025B03 5, mRNA sequence.
 ACCESSION BM448527
 VERSION BM448527.1 GI:19854099
 KEYWORDS EST.
 SOURCE Dunaliella salina.
 ORGANISM Dunaliella salina.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Dunaliellaceae; Dunaliella.
 REFERENCE 1 (bases 1 to 565)
 AUTHORS Cushman,J.C.
 TITLE An expressed sequence tag database for the halotolerant green alga,
 Dunaliella salina
 JOURNAL Unpublished (2002)
 COMMENT Contact: Cushman JC
 Department of Biochemistry
 University of Nevada
 MS200, Reno, NV 89557-0014, USA
 Tel: 775-784-1918
 Fax: 775-784-1650
 Email: jcushman@unr.edu
 PCR PRIMERS
 FORWARD: T3 20mer
 BACKWARD: T7 21mer
 Plate: 025 row: B column: 03
 Seq primer: T3 20mer
 High quality sequence stop: 565.
 Location/Qualifiers
 source 1..565
 /organism="Dunaliella salina"
 /db_xref="taxon:3046"
 /clone="DS4025B03"
 /clone_lib="An expressed sequence tag database for the
 halotolerant green alga, Dunaliella salina"
 /tissue_type="Cells, which was adapted in 2.5M NaCl via a
 incremental series from 1.7 to 2.0 to 2.25 to 2.5 M NaCl,
 were exposed to 3.4 M NaCl for 5 hours"
 /cell_type="Green"
 /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site.1:
 EcoRI; Site.2: XhoI. Library construction was performed
 according to Stratagene's recommended protocol for the
 Lambda UniZapXR vector and cDNA synthesis kit."
 BASE COUNT 131 a 163 c 179 g 92 t
 ORIGIN

Alignment Scores:
 Pred. No.: 5 86 Length: 565
 Score: 81.00 Matches: 36
 Percent Similarity: 41.33% Conservative: 26
 Best Local Similarity: 24.00% Mismatches: 54
 Query Match: 9.41% Indels: 34
 DB: 13 Gaps: 5

US-09-674-779b-2 (1-172) x BM448527 (1-565)

OY 9 AlaAlaAlaAlaAlaLeuSerValLeuThrPheMetThrGlyCysAlaAsnLysSerThr 28
 Db 11 GCAGAGCCGATGAAGCTTAAACATCTTACCCGCCACCTGGCTCCCAAGAAAGACCTGGAG 70
 OY 29 SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGly 48

Db 70 ----- 70
 OY 49 ValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLys---LeuPro 67
 Db 71 -----ATCATATATAGACGCAAGCTGCGGCTCTTACGACAAAGCGCTGTCC 118
 OY 68 SerLeuValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaIleLeu 87
 Db 119 GCAGAGCTGGAGGCGCAGGCCCTTGGCCACGACTTCAGAGGATTAATTTTCAAGATCATG 178
 OY 88 -----AspGluGlnAla-----GlnPheLeuThrThrAsnGlnThrAla 100
 Db 179 GCAGACAGACAGACAGACAGAGGTTTCGCCATGAAGAGGAGGTGCTCACCACCAACGCTGTG 238
 OY 101 ArgValLeuValAlaIleLysThrAspGluArgGlySerArgGluTyrAsnMetSerLeu 120
 Db 239 CGCCTCTCATGTCC-----CCCGAGACCAAGCTTCAGGGGTGATGCGCGCTGAT 292
 OY 121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuGlyLysGlyLeuAsnGlnAla 140
 Db 293 GCGGACGCGAGCGGCGCAAGAGCTGTGCGCGCTGCATCGTCTCTGATCTGTCTGCTG 352
 OY 141 SerValGluIleIleSerPheGlyGluGlu 150
 Db 353 AACCTGTGATCGTGAAGAAGGCGCAGCAG 382

RESULT 26
 BE372739 620 bp mRNA linear EST 21-JUL-2000
 LOCUS 601224250F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3582660 5',
 DEFINITION mRNA sequence.
 ACCESSION BE372739
 VERSION BE372739.1 GI:9318102
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 620)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (ULNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ULNL at:
 http://image.jnl.gov
 Plate: LAM8738 row: c column: 13
 High quality sequence stop: 598.
 Location/Qualifiers
 source 1..620
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3582660"
 /clone_lib="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary. Vector: pCMV-SPORT6; Site.1: SalI;
 Site.2: NotI. Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 137 a 179 c 175 g 129 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6 84 Length: 620
 Score: 81.00 Matches: 41

QY 111 gGlySerArgGlu 115
 Db 385 AGCCAGAGGAG 397

RESULT 30
 BM475133/c 1010 bp mRNA linear EST 05-FEB-2002
 LOCUS AGENCOURT_6477419 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559273
 DEFINITION 5', mRNA sequence.
 ACCESSION BM475133
 VERSION BM475133.1 GI:18524175
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1010)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: AHC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM12283 row: 3 column: 10
 High quality sequence stop: 517.

FEATURES
 Source Location/Qualifiers
 1..1010
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5559273"
 /clone_lib="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
 oligo-dT primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 316 a 224 c 214 g 256 t

ORIGIN

Alignment Scores:
 Pred. No.: 20.7 Length: 1010
 Score: 80.00 Matches: 34
 Percent Similarity: 47.338 Conservative: 28
 Best Local Similarity: 25.958 Mismatches: 55
 Query Match: 9.298 Indels: 14
 DB: 13 Gaps: 5

US-09-674-779b-2 (1-172) x BM475133 (1-1010)

QY 30 GluValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGlyVal 49
 Db 671 CAGGTGATGTTGTGAGAGAGATCTGAAATAGTGATATTTCTCTGCAGAGATT 612

QY 50 AlaProLeuValAspAsnAspGluThrValValAlaLeuAlaSerLysLeuProSerLeu 69
 Db 611 -----TTGGTCTGAATCAACCCAGTTTAAAACTCCGCGGCAAGATTATCAAGAATG 558

QY 70 ValTyrPheAspPheAspSerAspGluIleLysProGlnAlaAlaAlaIleLeuAspGlu 89
 Db 557 GTGTCATCTTCATCAGAGAAAGAACCATCAGCATCTGCGGCAAGTTCTTCAACAAT 498

QY 90 GlnAlaGlnPheLeuThrAsnGlnThrAlaValAlaValAlaGlnIleThrAsp 109
 Db 497 GCGACCTCTTCTGGGAAATCTCGCTCTCTCG-----GCCTTTCAC----- 456

QY 110 GluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaArg 129
 Db 455 ---CAGCAAGCAATGATTAATTCCTTGGACCTTG-----GCCCTTCCAAATTAAGA 411

QY 130 AsnTyrLeuLeuGlyGlyGlyLeuAsnGlnAlaSerValGlu---IleIleSerPheGly 148
 Db 410 ACTTACATCATCTGTAAAGGTCCCTTCAGCATCTCTCAAAATTCATCTGTACGTTGGC 351

QY 149 GluGluArgProIleAlaPheGlyThrAsnGlu 159
 Db 350 AATGTTATTCCTGTTCTCTCTGTTATTCATCA 318

RESULT 31
 BI670158/c 1131 bp mRNA linear EST 12-SEP-2001
 LOCUS 603294002F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:531336 5',
 DEFINITION mRNA sequence.
 ACCESSION BI670158
 VERSION BI670158.1 GI:15584391
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1131)
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM11794 row: 3 column: 01
 High quality sequence stop: 4.

FEATURES
 Source Location/Qualifiers
 1..1131
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:531336"
 /clone_lib="NIH_MGC_96"
 /tissue_type="hypothalamus"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pBluescriptR (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgaag
); Oligo-dT primed using primer 5'-TTTCTTTTCTTTTCTTTT-3',
 size-selected for average insert size 2.3 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 242 a 277 c 359 g 253 t

ORIGIN

Alignment Scores:
 Pred. No.: 25 Length: 1131
 Score: 80.00 Matches: 35
 Percent Similarity: 41.748 Conservative: 13
 Best Local Similarity: 30.438 Mismatches: 49
 Query Match: 9.298 Indels: 18
 DB: 13 Gaps: 4

US-09-674-779b-2 (1-172) x BI670158 (1-1131)

QY 29 SerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGly 48
 Db 497 GCGACCTCTTCTGGGAAATCTCGCTCTCTCG-----GCCTTTCAC----- 456

Db 336 TCCCAATACATG-----TCCCGGGGGGTCCAGGAGAGTCTTACACTCT 289

Qy 49 ValAlaProLeuValAspAsnAspGluThrVal-----LysAlaLeuAlaSerLysLeu 66

Db 288 GTGACATTATTCATTAAGACAGAAAGACGACCACTCTGACAGACCTACTCTCTG 229

Qy 67 ProSerLeuValTyrPheAspPheAspSerAspGluLeuLysProGlnAlaAlaAla 86

Db 228 CCCCTGCTCCCTAGACCTCTTACACCGGCTCTACTGTACAGACGACGACGACCTT 169

Qy 87 LeuAspGluGlnAlaGlnPheLeuThrThraSngInThrAlaArgValLeuAlaGly 106

Db 168 TCGATGAGTCTAGACCTCTCTCATAGTCCGCGCAGCCGACAGCAGTC----- 121

Qy 107 HisThrAspGluArgGlySerArgGluTyrAsnMetSerLeuGlyGluArgAlaVal 126

Db 120 -----CATGATCTCGAAGCTCTCGCTTCA-----CGAGGTGTG 85

Qy 127 AlAlaAlaArgAsnTyrLeuLeuGlyLysGlyIleAsnGlnAlaSer 141

Db 84 GCGCTCCCTCCTCCTCCGCTCGAGCATGAGTGTGCGCTGATCG 40

RESULT 32 385 bp DNA linear GSS 25-MAY-2001

LOCUS A0847628 Leishmania major FV1 random genomic library

DEFINITION Leishmania major genomic clone LMAJFV1_lm35c11 5', DNA sequence.

ACCESSION A0847628

VERSION A0847628.1 GI:6052276

KEYWORDS GSS.

SOURCE Leishmania major

ORGANISM Leishmania major

Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 385)

AUTHORS Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Blistein, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, H., R., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R. and Beverley, S.M. A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA microarrays and expression profiling

TITLE Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)

JOURNAL 21192569

MEDLINE Contact: Akopyants, NS / Beverley, SM

COMMENT Washington University School of Medicine

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

If using this information please cite:

N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing' and the Washington University Genome Sequencing Center for information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu)

Seq primer: -40RP from Gibco

Class: shotgun

High quality sequence stop: 365.

Location/Qualifiers

1..385

/organism="Leishmania major"

/strain="Friedlin strain VI"

/db_xref="taxon:5664"

/clone="LMAJFV1_lm35c11"

/clone_lib="Leishmania major FV1 random genomic library"

/lab_host="TOP10 (Invitrogen)"

/note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV; Genomic DNA was isolated from stationary phase cells. For this library, DNA was sheared to give a tight size distribution of 1-1.5kb fragments, blunt-ended with T4 DNA polymerase, dephosphorylated with Shrimp Alkaline Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 97 a 120 c 115 g 53 t

ORIGIN

Alignment Scores:

Pred. No.: 4.83 Length: 385

Score: 79.50 Matches: 34

Percent Similarity: 38.89% Conservative: 15

Best Local Similarity: 26.98% Mismatches: 44

Query Match: 9.23% Indels: 33

DB: 17 Gaps: 5

US-09-674-779b-2 (1-172) x A0847628 (1-385)

Qy 34 AlAlaProAsnAlaProThrGlyTyrThrGlyValIleTyrThrGlyValAlaProLeuVal 53

Db 8 GCGCCAGGACCGCCGCGGAGGTGACCGGACGCTCATCACCCTGACGGAATCTGCT 67

Qy 54 AspAsnAspGluThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAsp 73

Db 68 -----GCTTGAAGCAGAGGCGCGGTGAGAGGAGGAGGAGC 100

Qy 74 PheAspSerAspGluThrValLysProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPhe 93

Db 101 TTTGATGCCGCGGAGGACTCAAGCG-----CAGATC 133

Qy 94 LeuThrThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGluArgLysSer 113

Db 134 ATCCACATC-----GAGAACGCCGAG 154

Qy 114 ArgGluTyr---AsnMetSerLeuGlyGluArgAlaValAlaAlaArgAsnTyrLeu 132

Db 155 CGCGAGAGTTCACACGCTCGACACTAGAAAGACGCGCGAGTGTGGAGAACATCAAAA 214

Qy 133 LeuGlyLysGlyIleAsnGlnAlaSerValGluIleLeuSerPheGlyGluGluArgPro 152

Db 215 CTGACCAAGAGGCTC---AAGCTGCGTATTGAGACCTCCGAGTGGCAACAGAGAGATC 271

Qy 153 IleAlaPheGlyThrAsn 158

Db 272 GCGCGTCCCGCCACCAAC 289

RESULT 33 480 bp mRNA linear EST 21-JUN-2000

BE135571 u953a03.y1 Barstead bowel MFLRB9 Mus musculus cDNA clone

LOCUS IMAGE:1546060 5', similar to TR:060817 060817 NASCENT

DEFINITION POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE ;, mRNA sequence.

ACCESSION BE135571

VERSION BE135571.1 GI:8598090

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 480)

AUTHORS Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus.

Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810


```

Db      304 CCAACCTTCATGACTACTGCTACACCAAGTAATGGCAGTACC---ACTATTAT 360
QY      47 ThrGlyValAlaProLeu-----ValAspAsnAspGluThrValLys 60
Db      361 ACCACTACTTCCCTATCGATACACACTGTCACCGAGGTGATCTGATGCAACCATCT 420
QY      61 AlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLys 80
Db      421 GATGTAAATATCACACACCAACCTTTC----- 447
QY      81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrAsnGlnThrAla 100
Db      448 -----TCCTGACTACTGCTACACCAAGTAACGGTGTCTTACTACCGTCTACACCCACC 501
QY      101 ArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlyTyrAsnMetSerLeu 120
Db      502 ACTTGTCTCTATCGATACACACTACCACTGAAGTGCATCTAATGATGATACACTTCAAG 561
QY      121 GlyGluArgArgAlaValAlaValArgAsnTyrLeuLeuGlyLysGlyIleAsnGln--- 139
Db      562 TCGGAGATCGATGAATTTCTACACAGAACCTCATTTGCTACAGGTACACAAACAAAGAC 621
QY      140 -----AlaSerValGluIleIleSerPheGlyGlu-----GluArgPro 152
Db      622 CAAGTCTAGTGTGAGTAACAACCGCTTCTCAAAATGCTGAAGTCTCATCAGCCGAC 681
QY      153 IleAlaPheGlyThrAsnGluGluAlaTrpSerGlnAsnArgArgAlaGluLeuSer 171
Db      662 ATCGCTATCACTTGTGATGTGTCAATCGAAAGTGAAACGACAGACAAACAAAGC 738

RESULT 35
W20805      460 bp      mRNA      linear      EST 10-SEP-1996
DEFINITION      mbJ01c06.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:336778 5' similar to PIR:S49326 S49326 Nascent polypeptide
ACCESSION      W20805
VERSION      W20805.1 GI:1297704
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      1 (bases 1 to 460)
      Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
      Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
      Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
      Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
      Waterson,R.
TITLE      The WashU-HMT Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      Contact: Maria M/Mouse EST Project
      WashU-HMT Mouse EST Project
      Washington University School of MedicineP
      4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
      Tel: 314 286 1800
      Fax: 314 286 1810
      Email: mouseest@watson.wustl.edu
      This clone is available royalty-free through LNL; contact the
      IMAGE Consortium (info@image.llnl.gov) for further information.
      MGI:218178
      Seq primer: EMP1mer
      High quality sequence stop: 454.

FEATURES
source      1..460
      location/Qualifiers
      1..460
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="IMAGE:336778"
      /clone_lib="Soares mouse p3NMF19.5"
      /dev_stage="19.5 dpc total fetus"
      /lab_host="DH10B (ampicillin resistant)"
      /note="Vector: pT7T3D (Pharmacia) with a modified

```

```

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCCTACCAATCTGAAGTGGAGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University).

BASE COUNT      133 a      119 c      122 g      86 t
ORIGIN

Alignment Scores:
Pred. No.:      7.54      Length:      460
Score:      79.00      Matches:      39
Percent Similarity:      35.26%      Conservative:      16
Best Local Similarity:      25.00%      Mismatches:      57
Query Match:      9.18%      Indels:      44
DB:      14      Gaps:      5

US-09-674-779b-2 (1-172) x W20805 (1-460)

QY      26 LysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThrGlyValIle 45
Db      24 CGTTCTCTTTCTGCGCCGATCTGTTGCGTATCTCCGACAAATGCCCGT----- 77
QY      46 TyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeuAlaSerLys 65
Db      78 -----GAAAGCCACAAACCGTCCCTCTACAGACGAGG 113
QY      66 LeuProSerLeuValTyr-----PheAspPheAspSerAspGluLys 80
Db      114 TTGGCCACAGCCTCAGGCTGAGACAGATCGGAGACAGAGCTGACAGTATGATGACGTA 173
QY      81 ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrAsnGlnThrAla 100
Db      174 CCAAG-----CTCGAGAAACAGACTCCACACAGCGCCAGCAGCAAGCC 221
QY      101 ArgValLeuValAlaGlyHisThrAsp----- 109
Db      222 CAGCTGGCAGCCCGCAGCAGAGATGGAAGAACTGTAGTAAGCAAGCAGACTGCA 281
QY      110 -----GluArgGlySerArgGlyTyrAsnMetSerLeuGlyGluArgArgAlaValAlaVal 128
Db      282 AGTGAAGAAGAGCAAGCAAGAGCTATGCTCAAACTGGGCTTGCACAGGTACAGGGGTT 341
QY      129 ArgAsnTyrLeuLeuGlyLysGlyIleAsn----- 138
Db      342 ACGAGATCACTATCCGAAATCTAAATAATATCTCTTGTCTATCAAAACCCGATGTC 401
QY      139 -----GlnAlaSerValGluIleIleSerPheGlyGluGluArg 151
Db      402 TACAAGAGCCCGACCTTCAGACACCTACATATGTGTGGGGAACCAAGA 449

RESULT 36
BG261745      1083 bp      mRNA      linear      EST 13-FEB-2001
LOCUS      BG261745
DEFINITION      60237532F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4481142 5',
IMAGE sequence.
ACCESSION      BG261745
VERSION      BG261745.1 GI:12771561
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      1 (bases 1 to 1083)
      NIH-MGC http://mgc.ncl.nih.gov/
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished (1999)
      Contact: Robert Strausberg, Ph.D.
      Email: cgabs-r@mail.nih.gov

```

Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: LLM10315 row: p column: 07
 High quality sequence stop: 641.
 Location/Qualifiers

FEATURES

source

1.1083
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4481142"
 /clone_id="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; cloned unidirectionally; oligo-dt primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."
 BASE COUNT 295 a 276 c 285 g 227 t
 ORIGIN

Alignment Scores:

Pred. No.: 31.3 Length: 1083
 Score: 79.00 Matches: 30
 Percent Similarity: 44.04% Conservative: 18
 Best Local Similarity: 27.52% Mismatches: 38
 Query Match: 9.18% Indels: 24
 DB: 12 Gaps: 4

US-09-674-779b-2 (1-172) x BG261745 (1-1083)

QY 55 AsnAspGluThrValValAlaLeuAlaSerLysLeuProSer-----LeuValTyr 71
 Db 485 AGCAATGCAATATCCAGAAATATCCAGAGATCTCTCTGTAACAGTGTGTAT 544
 QY 72 PheAspPheAspSerAspGluLeuLysProGln----- 82
 Db 545 TTGGACTTC-CATTCAGATCATTCATCCCATGAGATTTTAAAGGACCTCCATCAAC 603
 QY 83 -----AlaAlaAlaLeuAspGluGlnAlaGlnPhe 93
 Db 604 TGAGAGAGGTTAACTGTCGAAAAAACGGCATCGAGTGTATTTGGCCAGCATTCG--TTC 660
 QY 94 LeuThrThrAsnGlnThrAlaArgValLeu--ValAlaGlyHisThrAspGluArgGly 112
 Db 661 CAAAGGGTAGAGAAACCCCTGGAGAACCTGGACTGTCTGGCCACCGGATTCACAGGGGT 720
 QY 113 SerArgGluTyrAsnMetSerLeuGlyGluArgAlaValAlaValArgAsnTyrLeu 132
 Db 721 GCCCAAAATATGCTCTCAATCTCTGGAAGGCAAGGCGTAAAGAGGTATTAACCCACG 780
 QY 133 LeuGlyLysGlyLysGlnAlaSer 141
 Db 781 GTTGTGGGGGGGTCTCCCAAGTTTCC 807

RESULT 37
 BE261626 503 bp mRNA linear EST 26-OCT-2000
 LOCUS 601149316F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3502299 5',
 DEFINITION mRNA sequence.
 BE261626
 ACCESSION BE261626.1 GI:9133944
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 503)
 AUTHORS NIH-MGC <http://mgs.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at: image.llnl.gov
 Plate: LLM173 row: o column: 04
 High quality sequence start: 3
 High quality sequence stop: 502.
 Location/Qualifiers

FEATURES

source

1.503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3502299"
 /clone_id="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dt priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."
 BASE COUNT 105 a 134 c 141 g 123 t
 ORIGIN

Alignment Scores:

Pred. No.: 10.1 Length: 503
 Score: 78.50 Matches: 26
 Percent Similarity: 42.45% Conservative: 33
 Best Local Similarity: 18.71% Mismatches: 61
 Query Match: 9.12% Indels: 19
 DB: 10 Gaps: 5

US-09-674-779b-2 (1-172) x BE261626 (1-503)

QY 23 CysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThr 42
 Db 55 TGCTCTCGCACACACCCACTTCATTTGGCCAAACACCGACCCCGAATTCTCC 114
 QY 43 GlyValIleTyrThrGlyValAlaProLeuValAspAsnAspGluThrValLysAlaLeu 62
 Db 115 GCMAATTCCTGTCCGGCTAGGTCCAGATATTGAGGTGATGGATTCTGTGTGT 174
 QY 63 AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluLys----- 80
 Db 175 GAGACCAAC-----GCCATTGCTTACTAT--GTGACCAATGAGGAGCTGGCGGCAAGT 225
 QY 81 ---ProGlnAlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThr 99
 Db 226 ACTCCAGAGCAGACAGCCAGCGGTGGTGCAGTGGTGTGCTGATTCGACCTATAGT 285
 QY 100 AlaArgValLeuValAlaGlyHisThrAspGluArgGlySerArgGlyTyrAsnMetSer 119
 Db 286 GCCCCAGCCAGTACTGTGTGTGTCCTCCACCTTGGGCAATGACACCAACAAACAG 345
 QY 120 LeuGlyGlu-----ArgArgAlaValAlaVal----- 128
 Db 346 GCCACTGAGATGAAAGAGAGAGAGTGAAGCGAATTTCTGGGGCTGCTGCTTACTTG 405
 QY 129 -----ArgAsnTyrLeuLeuGlyLysGlyLysAsnGlnAlaSerValGluIleIle 145
 Db 406 AAGACGAGACTTTTGTGTGGGGAACGATGACATTTGCTGACATCAGATTGTC 462
 RESULT 38
 D34784 360 bp mRNA linear EST 08-AUG-1994
 LOCUS D34784

DB	340	TCGGATGCTTCATC	354	
RESULT 39				
AQ954996/c				
LOCUS				
DEFINITION	AQ954996	470 bp	DNA	linear
ACCESSION	nb00076M14f	CUG1 Rice BAC library (EcoRI)	Oryza sativa genomic	
VERSION	AQ954996	clone OSJNB0076M14f, DNA sequence.		
KEYWORDS	AQ954996.1	GI:6778262		
SOURCE	GSS.			
ORGANISM	Oryza sativa.			
	Oryza sativa			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
	Ehrhartoideae; Oryzaceae; Oryza.			
	1 (bases 1 to 470)			
REFERENCE	Wing, R.A. and Dean, R.A.			
AUTHORS	A BAC End Sequencing Framework to Sequence the Rice Genome			
TITLE	Unpublished (1998)			
JOURNAL	Contact: Wing RA			
COMMENT	Clemson University Genomics Institute			
	100 Jordan Hall, Clemson, SC 29634, USA			
	Tel: 864 656 7288			
	Fax: 864 656 4293			
	Email: rwing@clemson.edu			
	Seq primer: GTAAACGACGCGCAGTG			
	Class: BAC ends			
	High quality sequence start: 73			
	High quality sequence stop: 332.			
FEATURES	Location/Qualifiers			
SOURCE	1..470			
	/organism="Oryza sativa"			
	/strain="Japonica"			
	/cultivar="Nipponbare"			
	/db_xref="taxon:4530"			
	/clone="OSJNB0076M14f"			
	/clone_lib="CUG1 Rice BAC library (EcoRI)"			
	/russ_type="leaf"			
	/lab_host="E. coli DH10B"			
	/note="Vector: pBACindigo; Site.1: EcoRI; Site.2: EcoRI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."			
BASE COUNT	83 a	142 c	143 g	99 t
ORIGIN				3 others
Alignment Scores:				
Pred. No.:	10.5	Length:	470	
Score:	78.00	Matches:	29	
Percent Similarity:	46.15%	Conservative:	19	
Best Local Similarity:	27.88%	Mismatches:	52	
Query Match:	9.06%	Indels:	5	
DB:	17	Gaps:	2	

US-09-674-779b-2 (1-172) x AQ95496 (1-470)

QY 23 CysAlaAsnLysSerThrSerGlnValMetValAlaProAsnAlaProThrGlyTyrThr 42
 DB 339 TGTACCCGCTCTCTCTGCTGAGCGAGAGCCCGGAGT-----ACCGGCGAGTGC 286
 QY 43 GlyValIleTyrThrGlyValAlaProLeuValAspAsnGluThrValLysAlaLeu 62
 DB 285 GCGGTGGCCGCTACGCGGGG-GAGGACGTGCTGACGACGACGACGACATGCTCCGTC 227
 QY 63 AlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAspGluIleLysProGln 82
 DB 226 GCGGACGCTGTGTGCAGC-----TACCTCTCTCTGACATGACGACGCTGAGATATAC 173
 QY 83 AlaAlaAlaIleLeuAspGluGlnAlaGlnPheLeuThrThrAsnGlnThrAlaArgVal 102
 DB 172 ACCGCGAGGTGGCTTCCACTCAAGACGACAAACCGCGCGGAGTGTGGCGGCGCA 113
 QY 103 LeuValAlaGlyHisThrAspGluArgGlySerArgGlyTyrAsnMetSerLeuGlyGlu 122
 DB 112 CTGGAAAGCCGGAAGAGCAGGAGCAAGCCGCGGGAAGCCGCTCATCTTGGCAAG 53
 QY 123 ArgArgAlaVal 126
 DB 52 AGGACAGTGTCTC 41

RESULT 40
 BM491955 676 bp mRNA linear EST 07-FEB-2002
 LOCUS p2n.pk007.o23 Normalized Chicken Pituitary/Hypothalamus/Pineal
 DEFINITION Library (p2n) Gallus gallus CDNA clone p2n.pk007.o23 5' similar
 to results pending, mRNA sequence.
 ACCESION BM491955
 VERSION BM491955.1 GI:18612886
 KEYWORDS EST.
 SOURCE chicken.
 ORGANISM Gallus gallus
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 676)
 AUTHORS Porter, T.E. and Cogburn, L.A.
 TITLE ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal CDNA
 library, USDA/IRAFs Animal Genome Project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Larry A. Cogburn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: cogburnudel.edu, www.chickest.udel.edu.

FEATURES
 source
 1..676
 /organism="Gallus gallus"
 /strain="Commercial broiler chickens"
 /db_xref="taxon:9031"
 /clone="p2n.pk007.o23"
 /clone_lib="Normalized Chicken
 Pituitary/Hypothalamus/Pineal Library (p2n)"
 /sex="Male and Female"
 /tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
 /dev_stage="Embryonic (d12,d14,d19); post-hatch (1,3,5,7,9
 weeks)"
 /lab_host="E. coli EMD10B"
 /note="Vector: PCMVSPORT6; Library made from equivalent
 pools of total RNA isolated from each tissue at different
 ages. Single pass sequencing from 5'-end"
 BASE COUNT 99 a 230 c 227 g 105 t 15 others
 ORIGIN

Alignment Scores: 19.2 Length: 676
 Pred. NO.: 78.00 Matches: 40
 Score:

Percent Similarity: 34.91% Conservative: 19
 Best Local Similarity: 23.67% Matches: 74
 Query Match: 9.06% Indels: 36
 DB: 13 gaps: 6

US-09-674-779b-2 (1-172) x BM491955 (1-676)

QY 8 AlaAlaAlaAlaAlaAlaSerValIleuThrPheMetThrGlyCysAlaAsnLysSer 27
 DB 41 GCGCGCGCGCGCGCGCGCGCGCGCGCATACCTTGGCGGGCGCGCGCGAGACCGG 100
 QY 28 ThrSerGlnValMetValAlaProAsnAlaProThr-----Gly 40
 DB 101 CAGCAGCAGACCTGGCGGGCGCGCGCGCGCGCGCGCGTGGCGGTCGCGCGGCA 160
 QY 41 TyrThrGlyValIleTyrThrGlyValAlaProLeu-----ValAspAsnAspGlu 57
 DB 161 CCGTGGGTCCGCGAGCCCGCGCGCGCGCGCGCGCGCATTCGCGGCTGTCCACTGGCC 220
 QY 58 ThrValLysAlaLeuAlaSerLysLeuProSerLeuValTyrPheAspPheAspSerAsp 77
 DB 221 GCGATCCGCGCGGAGATCGCGCGCTTGCAGTGC----- 253
 QY 78 GluIleLysProGlnAlaAlaAlaIleLeuAsp-----GluGlnAlaGlnPheLeuThr 95
 DB 254 --GTGCACCCCAACATCTACGCGCATCTACGCGCATCTACGCGCGCGTGGAGACCTGGCG 310
 QY 96 ThrAsnGlnThrAlaArgValLeuValAlaGlyHisThrAspGlu--ArgLysSerArg 114
 DB 311 CTGCAGAGCCGATCCGCGAGCACGTCATCTCCATCGAGGATTCCTTTGGAAACAGCCAG 370
 QY 115 GluTyrAsnMetSer-----LeuGly 121
 DB 371 GAATGGAACCCCGACCCGCTGCTGAGCTGAAGGTGGGACATCGGGGAACCTGTCC 430
 QY 122 GluArgArgAlaValAlaValAlaArgAsnTyrLeuGlnTyrLysGlyIleAsnAlaSer 141
 DB 431 AGCGCAAGTGGCGCTGTGTGCACCGGCTGACCGGACCTATGTGTGACGAGAGTCT 490
 QY 142 ValGluIleIleSerPheGlyGluGlu 150
 DB 491 CCGGAGGTGGCGCATTTAAGAGAG 517

Search completed: July 6, 2003, 13:11:03
 Job time : 1417 secs